

International Speech Communication Association

Proceedings of ISCA Tutorial and Research Workshop

on

# **Experimental Linguistics**

25-27 August 2008, Athens, Greece



Edited by Antonis Botinis



University of Athens

**ISCA**  
International Speech Communication Association  
Proceedings of ISCA Tutorial and Research Workshop On  
**Experimental Linguistics**  
**ExLing 2008**  
25-27 August 2008, Athens, Greece  
Edited by Antonis Botinis

**University of Athens**

ISCA Experimental Linguistics ExLing 2008  
Antonis Botinis, editor

ISBN: 978-960-466-020-9

Copyright 2008  
ISCA and the University of Athens

## **Forword**

---

This volume includes the proceedings of the second ISCA (International Speech Communication Association) Tutorial and Research Workshop on Experimental Linguistics, ExLing 2008, held in Athens, Greece, 25-27 August 2008. This workshop is the third ISCA Workshop in Athens, the first one on Prosody, in September 1997, and the second one on Experimental Linguistics, in August 2006.

Two years ago, on the occasion of the 2006 ISCA Workshop on Experimental Linguistics, we gathered in Athens and discussed current developments in major fields of linguistics and the use of experimental methodologies in order to promote linguistic knowledge and enhance language applications. Many colleagues expressed their vivid interest in this initiative and embraced the idea for the repetition of the Workshop. We did promise to do our best and the set up of the present Workshop had a significant appeal to the international scientific community, which enthusiastically responded to our call.

Typical of the premises and goals of the Workshop, the volume includes a variety of experimental and interdisciplinary papers, ranging from speech production and language education to psycholinguistics and language technology. ISCA sponsored events provide an ideal environment for the promotion of new ideas and the development of research methodologies, the backbone of scientific investigations in theory and practice. Many papers are related to joint projects of research students and established researchers, which is a promising avenue in the spirit of the present Workshop.

All our colleagues from different parts of the world, including countries such as Brazil, Canada, Japan and the USA, are most welcomed and we hope for extensive scientific exchange and feedback as well as dynamic debates and the setting up of international collaboration. We also hope that everyone will have a good time and enjoy their stay in Greece.

We would like to thank all participants for their contributions as well as ISCA for the support and organisation of this event and the University of Athens for the publications of the proceedings. Special thanks also to our students at Athens University Department of Linguistics for their invaluable assistance in crucial organisation issues.

The organising committee

Aikaterini Bakakou-Orphanou  
Antonis Botinis  
Christoforos Charalambakis

## Contents

---

Discourse markers and L2 listening: do computers make a difference in L2 listening comprehension? Nuray Alagözlü	1
Word category and prosodic emphasis in dialog modules of speech technology applications Christina Alexandris	5
Saudi accented Arabic voice bank Mansour Alghamdi, Fayez Alhargan, Mohamed Alkanhal, Ashraf Alkhairy, Munir Eldesouki and Ammar Alenazi	9
A new Arabic stemming algorithm Eiman Tamah AlShammari and Jessica Lin	13
Neurolinguistic aspects of metaphor theory Georgia Andreou and Ioannis Galantomos	17
Prosodic variation in L2: a case of Germans speaking English Volha Anufryk, Matthias Jilka and Grzegorz Dogil	21
The influence of top-down expectations on the perception of syllable prominence Denis Arnold and Petra Wagner	25
Overspecification in action-oriented discourse: task importance affects the production of overspecifications and overspecifications increase identification efficiency in perception Anja Arts, Alfons Maes, Leo Noordman and Carel Jansen	29
Prosodic phrasing in German sentence production: optimal length vs. argument structure Petra Augurzky	33
Coarticulation in non-native speakers of English: /əIV/-sequences in non-proficient vs. proficient learners Henrike Baumotte and Grzegorz Dogil	37
Rhythm and stress intervals in Greek and Russian Antonis Botinis, Marios Fourakis and Olga Nikolaenkova	41
Investigations of speech segmentation: addressing the writing bias in language research Victor J. Boucher and Annie C. Gilbert	45



---

Two sources of voicing neutralization in Lithuanian Rebeka Campos-Astorkiza	49
Stress assignment in Brazilian Portuguese: a usage-based approach Maria Cantoni	53
The temporal structure of professional speaking styles in Brazilian Portuguese Luciana Castro and João Antônio de Moraes	57
Prosodic perception of sentence types in Greek Anthi Chaida	61
Classification by discriminant analysis of the energy in view of the detection of accentuated syllable in Standard Arabic Amina Chentir, Mhania Guerti and Daniel Hirst	65
The identification of the place of articulation in coda stops as a function of the preceding vowel: a cross-linguistic study Man-ni Chu, Carlos Gussenhoven and Roeland van Hout	69
Reading mathematical exercises: preliminary results Deolinda Correia, Isabel Hub Faria and Paula Luegi	73
Prosody in read aloud text: relation with information status, content type and boundary strength Hanny den Ouden and Carel van Wijk	77
Imperatives in European Portuguese: a perception approach Isabel Falé	81
Nasometric values for European Portuguese: preliminary results Isabel Falé and Isabel Hub Faria	85
Priming effect on word reading and recall Isabel Hub Faria and Paula Luegi	89
Formulaic expressions in language technology María Fernández-Parra	93
Continuation tunes in two central varieties of Italian: phonetic patterns and phonological issues Rosa Giordano	97
Model-based duration analysis on English natives and Thai learners Chatchawarn Hansakunbuntheung, Hiroaki Kato and Yoshinori Sagisaka	101

Lexicalization of natural actions and cross-linguistic stability Paul E. Hemeren, Sofia Kasviki and Barbara Gawronska	105
“Deep and raspy” or “high and squeaky”: a cross-linguistic study of voice perception and voice labelling Begoña Payá Herrero	109
The effect of focus on lexical tones in Vietnamese Stefanie Jannedy	113
MORPHEMIA: a semi-supervised algorithm for the segmentation of Modern Greek words into morphemes Constandinos Kalimeris and Stelios Bakamidis	117
The acquisition of temporal categorical perception by Japanese second language learners Naoko Kinoshita	121
Tongue movements and syllable onset complexity: ultrasound study Tanja Kocjancic	125
The prosodic and nonverbal deficiencies of French- and Finnish-speaking persons with Asperger Syndrome Mari Lehtinen	129
The effectiveness of auditory phonetic training on Greek native speakers’ perception and production of Southern British English vowels Angelos Lengeris	133
Phonetic convergence and language talent within native-nonnative interactions Natalie Lewandowski, Travis Wade and Grzegorz Dogil	137
A comparison of Taiwanese sign language and manually coded Chinese: word length and short-term memory capacity Hsiu-Tan Liu, Chin-Hsing Tseng and Chun-Jung Liu	141
The role of animacy in the production of Greek relative clauses Sofia Loui and Silvia P. Gennari	145
Acoustic model of stress in standard Greek and Greek dialects Anastassia Loukina	149
Using F2 transition parameters in distinguishing Persian affricates from homorganic consonants Zahra Mahmoodzade and Mahmoud Bijankhan	153

---

Intonation of parentheses in spontaneous French sentences Philippe Martin	157
Perception of consonant clusters in Japanese native speakers: influence of foreign language learning Hinako Masuda and Takayuki Arai	161
How are words reduced in spontaneous speech? Holger Mitterer	165
Phonological free variation in English: an empirical study Jose A. Mompean	169
Interaction of phonetics, phonology, and sociophonology – illustrated by the vowels of Standard Austrian German. Sylvia Moosmüller	173
PENS: a confidence parameter estimating the number of speakers Siham Ouamour, Mhania Guerti and Halim Sayoud	177
Phoneme classification using the Hartley Phase Spectrum Ioannis Paraskevas and Maria Rangoussi	181
The influence of music education and training on SLA Barbara Pastuszek-Lipińska	185
Rhythmic analysis and quantitative measures: the essence of rhythm as temporal patterning Michela Russo	189
Compensatory lengthening in Persian: the timing of non-modal phonation Vahid Sadeghi	193
Faster time-aligned phonetic transcriptions through partial automation Ben Serridge and Luciana Castro	197
The effects of the acoustic properties of second language vowel production on pronunciation evaluation Chris Sheppard	201
Frequency effects in language acquisition: a case study of plural forms in Brazilian Portuguese Thais Cristófaró Silva and Christina Abreu Gomes	205



Factors influencing perceptual attainment of Japanese geminate consonants by Korean learners of Japanese Mee Sonu	209
Receptive and productive skills of English /l/ and /r/ by Japanese college students in relation to their motivation Yuichi Todaka	213
Objective evaluation of second language learner's translation proficiency using statistical translation measures Hajime Tsubaki, Keiji Yasuda, Hirofumi Yamamoto and Yoshinori Sagisaka	217
Automatic labeling of prosody Agnieszka Wagner	221
Name dominance in spoken word recognition is (not) modulated by expectations: evidence from synonyms Andrea Weber and Alissa Melinger	225
Vocal stereotypes Melanie Weirich	229
Does uncertainty effect the case of exhaustive interpretation? Charlotte Wollermann and Bernhard Schröder	233
"Sounds like a rainbow" - sound-colour mappings in vowel perception Magdalena Wrembel and Karolina Rataj	237
<b>Supplement</b>	
Focus effects on syllable duration in Cypriot Greek Charalabos Themistocleous	241

# **Discourse markers and L2 listening: do computers make a difference in L2 listening comprehension?**

Nuray Alagözlü

Department of Foreign Language Teaching, Baskent University, Turkey

## **Abstract**

This study explores whether discourse markers play a role in L2 listening comprehension and whether software programs make a difference in L2 listening comprehension. Two groups of approximately 25 freshman students listened to two versions of three different talks with and without higher order discourse markers. The talks were delivered by a native speaker, a non-native speaker and by a software listening program (NaturalReader© 7.0). The scorings, compared using the Mann-Whitney U test, indicate that discourse markers play a significant role only in native speaker speech ( $p < 0,05$ ). It is seen that the highest mean of comprehension is found in the talks delivered by the native speaker. Computers ranked second in providing more comprehensive talks.

Key words: discourse markers, L2 listening comprehension

## **Introduction**

Discourse structuring devices, specifically discourse markers (DMs), are of importance in the organization of texts, both spoken and written. Defined as “sequentially dependent elements” which bracket “units of talk” (Shiffrin, 1992), or “information units” (Brown and Yule, 1985), they work at different levels of the organization of the talk (Shiffrin, 1992). DMs are particular kinds of cohesive devices. With cohesion in discourse, the direction of ideas becomes transparent and their sequence flows easily, which is necessary for the interpretation of a text.

“Macro DMs” are the signalling markers which mark the boundaries between episodes and moves in talks; more precisely, They help the listener or reader recognize information units and organize them in his mind. They trigger readers’ or listeners’ predictions and expectations about the text (as you remember, our story doesn’t end here, etc.). DMs which provide more opportunities for recognizing units at sentence and word level are called “micro DMs” (well, oh, because, etc.). The microstructure is associated fundamentally with a semantic, cohesive function holding between surface structure sentence (Chaudron and Richards, 1986).

In their study, Chaudron and Richards (1986) found that macro- DMs help more than micro-and macro-DMs together and more than micro- DMs alone in L2 learners’ understanding and recall of lectures. In several studies inspired by Chaudron and Richards (1986), the positive effects of the

presence of DMs in texts (Flowerdew and Tauroza, 1995; Williams, 1992; Zohreh and Rasekh, 2007) are shown although some indicated that DMs do not assist L2 listeners in comprehending English-medium lectures (Dunkel and Davis, 1994).

### **Problem**

ELT students in their English medium classes in the Turkish educational setting are usually exposed to various complicated lectures in L2 and need to decode them on their own. As revealed in informal interviews and observations, ELT students frequently want their teachers to retell or summarize what is told in their mother tongue at the end of the classes, which confirms they have difficulties and need to construct effective strategies in listening to lectures in L2. Listening efficiently in L2 is a difficult task that requires very careful observations and intelligent strategies on the part of the listener. It may be helpful for students to be provided with additional discourse-based strategies to aid their listening comprehension. Therefore, the probable effects of DMs on students' L2 listening comprehension are investigated in the present study as it is believed that such an awareness is an essential part of understanding spoken discourse. Listening to computer speech was also included in the design to see whether computer speech can offer comprehensive discourse where DMs play a role. This study aims to answer the following research questions: 1) Do DMs play a role in L2 listening comprehension? 2) Do software programs of listening make a difference in L2 listening comprehension?

### **Methodology**

As instruments to measure listening comprehension, two different versions of three texts [Artificial Life, The Cypriots, and Divorce and Kids] (Kıymazarslan et al., 2005) were used: one with very few micro markers and the other with macro markers in addition to few micro markers. The selected talks had very few DMs in nature. Macro DMs were added later and were revised for their information structure and authenticity. Two sections in the department were taken as the Groups A and B. At different times, Group A listened to the talks with few micro DMs while Group B listened to the version with macro markers read by the computer, by a native speaker or a non-native speaker. Thus, two groups of approximately 25 freshman students (50 in total; 47 females, 3 males) aged between 17 and 23 listened to two versions of three different talks with and without higher order DMs. The listeners heard the talks via two loudspeakers during their regular class hours in an electronic classroom. They listened to the talks twice. In the delivery of the talks via computer, a computer program "NaturalReader© 7.0" was used to convert written talks into spoken words. The written

comprehension scales (three different tests) contained five-item YES/NO questions ( $r=0,76$ ). A Mann-Whitney U Test was used to compare the means of the groups.

## Results

The presence of macro DMs was found to be statistically significant only in native speaker talk ( $p<0,05$ ). With macro DMs, a remarkable increase at comprehension levels was also seen in computer talk, but it was not statistically significant. The use of DMs in non-native speaker talk did not affect comprehension levels significantly. The highest mean score of comprehension was obtained from the talk with micro and macro DMs delivered by the native speaker. The software program ranked second in providing comprehensive talks when they contain micro and macro DMs. (Figure 1).

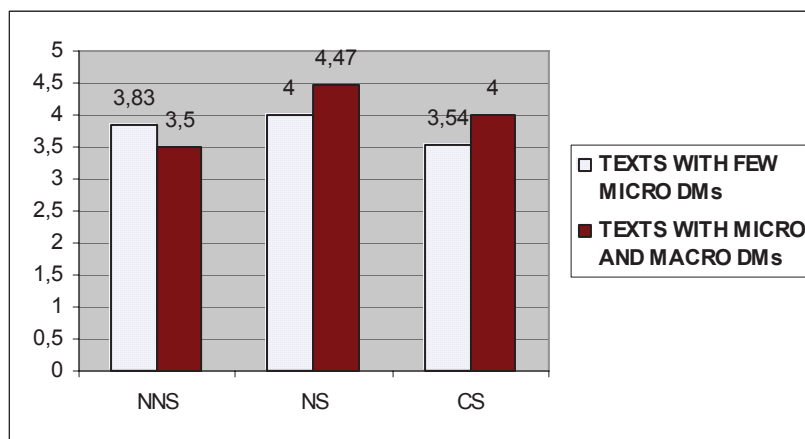


Figure 1. Comprehension Levels in L2 Listening. (NNS: Non-Native Speaker, NS: Native Speaker; CS Computer Speech).

## Conclusion

The findings of this research contributed to the idea that macro-DMs led to better recall of the lectures than lower level markers, micro-DMs. However, macro DMs are seen to play a significant role only in native speaker speech. This is quite reasonable since the listeners could make use of the native speaker's signals of major segments or transitions that probably add an organizational effect to the talks, which make the information more salient and meaningful. Native speaker's intonation and emphasis on major segments seem to help listeners.

Secondly, probably because of prosodic problems, the presence of DMs in computer speech did not play a significant role in comprehension. This suggests such software programs may work better in listening laboratories in case they should be improved in speech prosody as poor prosody can distract the listener and impede comprehension.

### **Acknowledgements**

I owe gratitude to Dr. Laurence Raw and to my freshman students (IDO Sections 1&2) for their valuable contribution to the study in the data collection process.

### **References**

- Brown, G and Yule, G. 1985. Discourse analysis. Cambridge: Cambridge University Press.
- Chaudron, C and Richards, J.C. 1986. The Effect of Discourse Markers on the Comprehension of Lectures. *Applied Linguistics*. Vol.7 No. 2. 113-127.
- Dunkel, P. A. And Davis, J. M. 1994. The effects of rhetorical signaling cues on the recall of English lecture information by ESL and ENL listeners. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 55-74). Cambridge: Cambridge University Press.
- Flowerdew, J. and Tauroza, S. 1995. The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, 17, 435-458.
- Kıymazarslan,V; Alagözlü, N. and Mirzayeva, N. 2005. Interactive Listening Booth. For Listening Practice. Seçkin Yayıncılık. Ankara.
- Shiffrin, D. 1992. *Approaches to Discourse*. Blackwell. Oxford.
- Williams, J. 1992. Planning, discourse marking and the comprehensibility of international teaching assistants. *TESOL Quarterly*, 25(4), 693- 708. Zohreh R. E. and Rasekh A.E. (2007) Discourse Markers in Academic Lectures. *Asian EFL Journal*. Volume 9, Issue 1.

# **Word category and prosodic emphasis in dialog modules of speech technology applications**

Christina Alexandris

Department of German, University of Athens, Greece, and  
Institute for Language and Speech Processing (ILSP) Athens, Greece

## **Abstract**

The prosodic behavior of word categories in Modern Greek is observed in relation to two keyword-based Speech Technology applications, namely for a Speech Recognition system in the healthcare domain for task-oriented dialogs intended for senior citizens and for the prosodic modelling of utterances produced by a Conversational Agent in a dialog system for consumer complaints. Word categories are evaluated and classified in the two applications, in respect to the types of words contained in each category, their relation to keywords and their effect on the user.

Key words: prosodic emphasis, semantic content, keywords, user-friendliness

## **System requirements and evaluation**

The behavior of word categories in Modern Greek is observed during the prosodic modelling of utterances recorded in a studio for the construction of the speech output produced by a Conversational Agent in a dialog system for consumer complaints (CitizenShield System, [www.polias.gr](http://www.polias.gr)), labelled “CSh System”, and for a dialog-module of a speech-recognition in a system for Senior Citizens which activates and monitors house appliances and the general room environment (SOPRANO Project, [www.soprano-ip.org](http://www.soprano-ip.org)), labelled “SPR System”. The acceptable prosody to the target user-groups, constituting the general public in the CSh-System and senior citizens in the SPR Project, is modelled with the use of emphasis as prosodic marker, placed on the appropriate elements of the utterance for achieving user-friendliness in man-machine communication in the sense of “accuracy” and “directness” (Hausser 2006) as well as (B) a sense of control to the user for the CSh System (Alexandris 2007, Nottas et al. 2007) and (C) familiarity for the SPR System as an additional feature, according to the SPR Project’s User Requirements. The evaluation of the prosodic markers in the SPR Project was primarily based on Interaction Aspects used for testing the performance of the CSh’s System, namely (1) the Utterance, (2) Functional and (3) Satisfaction Levels (Moeller 2005) and also due both to the complexity of the CSh’s System (a) user-group, a remarkable subset of which constitutes senior citizens, according to the CSh’s Project User Requirements (Work-Package 1) and the (b) relatively positive results of the evaluation of the systems prosodic modelling in the Evaluation Phase (Work-Package 7) (of a



50-Test Users Group, 60% Men and 40% Women, Average Age: 35 – 45, 80%, evaluated the Interface as “Very natural-sounding”, and 50% evaluated the System as “Seems to inspire very high credibility”). The prosodic modeling of the output in the SPR Project was subjected to an internal evaluation before the execution of user-tests to Senior Citizens.

In the present study we attempt to provide a differentiation between specific word categories in which prosodic emphasis does not determine their semantic content (I) and word categories whose semantic content may be determined by prosodic emphasis (II). In the first case, the semantic interpretation of the entire phrase or sentence may be determined by the type of element receiving prosodic emphasis, but the semantic content of the emphasized element itself is not effected. Based on the empirical data obtained both from the recorded corpora (ASR Component) and the evaluation of the spoken utterances produced by the Conversational Agent (Nottas et al. 2007) in the CSh and SPR Projects we attempt to go a step further and provide (1) an integration of the results of previous studies (Alexandris et al. 2005, Alexandris 2007) and to proceed to a (2) differentiation between specific word categories in relation to prosodic emphasis and their semantic content.

### **Prosody, word categories and semantic content**

The group of word categories whose semantic content may be determined by prosodic emphasis namely (1) spatial and temporal expressions, (2), a subgroup of quantifiers and numerals and (3) a sub-group of discourse particles identified as “politeness markers” (Alexandris and Fotinea, 2004) is classified as Category A or “Prosodically Determined” words. For spatial and temporal expressions, and for the subgroup of quantifiers and numerals, the presence of prosodic emphasis signalizes an indexical interpretation (“exactly”) as opposed to a vague (Schilder and Habel 2001), interpretation or a fixed expression (Alexandris et al. 2005), where in the latter cases, there is an absence of prosodic emphasis. For example, with prosodic emphasis there is an indexical interpretation of the spatial expression “dipla” as “along” in the sentence “the crack was exactly along (parallel) to the band in the packaging” as opposed to its vague interpretation as “next-to” in the same sentence. The same is observed for the temporal expression “oso” with its indexical interpretation as “for as long as” in the sentence “the array is created for as long as the loop is running” as opposed to its vague interpretation as “while” in the same sentence. Similarly, the numerical or quantificational expression “two” (“d'yo”) is used in its indexical and literal meaning when it receives prosodic emphasis in the sentence “wait for two minutes”, while, in the same sentence without prosodic emphasis, it is perceived as a fixed expression (“wait a moment”).

For discourse particles identified as “politeness markers”, the absence of prosodic emphasis signalizes them as politeness markers, while with the presence of prosodic emphasis they only have the property of discourse particles. Thus, absence of prosodic emphasis in the discourse particles “Tell me ('pite mou)” and “Mabey” ('mipos) signalizes positive politeness and friendliness towards the User in the following utterances produced by the Conversational Agents for the CSh and SPR Systems respectively: Tell me ('pite mou), what is the product (Preferred utterance by CSh Users), Mabey ('mipos) you want me to check the kitchen? (SPR)

The group of word categories where prosodic emphasis may emphasize or intensify, but may not determine the semantic content, is classified as Category B or “Prosodically Sensitive” words. This group involves (1) adjectives expressing quality and (2) adverbs expressing mode perceptible to the senses, used in a literal, non-metaphorical way. For example, prosodic emphasis on the adjective “round” (“strogi 'lo”), in the sentence “It was in a round box” (CSh Project) signalizes the meaning “truly/par excellence” round”. Similarly, prosodic emphasis on the adverb “upside down” (“an 'apoda”) for example, in the sentence “I turned it upside down” (CSh Project) signalizes the meaning “completely upside down”.

Both Category A and Category B type prosodic emphasis may be used in both the CSh and the SPR Projects, for the (a) correct interpretation of Speaker Input in the respective Automatic Speech Recognition (ASR) Modules and (b) for achieving user-friendliness in man-machine communication in the sense of “accuracy” and “directness” (Hausser, 2006) towards the user in the Conversational Agent’s spoken output.

The rest of the word categories that are not effected by prosodic emphasis in respect to their semantic content are classified as Category C or “Prosodically Independent” words. The presence or absence of prosodic emphasis on words of Category C only effects the semantic interpretation of the entire phrase or sentence in which they belong. A significant percentage of these words are nouns or verbs and they may constitute sublanguage-specific keywords. Prosodic emphasis on keywords focuses on the basic content of the utterance, for example, whether it is an action in question, in the case of a verb, or a specific object in question, in the case of a noun. Prosodic emphasis on the word elements of Category C, words is sentence-dependent and highly sublanguage- and application-specific. Prosodic emphasis on elements of Category C in the CSh Project is used both for (a) determining the basic content of the Speaker’s input, (b) for directing the Speaker’s input towards a keyword-specific answer, as well as (c) for achieving accuracy and directness in the Conversational Agent’s output. For example, in the sentence “Please tell us any additional information you wish about the product or about your transaction” (CSh), the keywords

“additional”, “product” and “transaction” receive prosodic emphasis for clarity towards the Users and simultaneously direct towards obtaining a respective keyword-specific answer, in this case “product-type” and “transaction-type”. Similarly, a “Yes/No” Answer is requested with the use of prosodic emphasis either on “check” or on “thermostat” in the question “Shall I check the thermostat?” in the SPR Project.

### **Systematic use of prosodic emphasis features**

In contrast to both A and B word categories, or “Prosodically Determined” and “Prosodically Sensitive” words, whose plus or minus ( $\pm$ ) prosodic emphasis features can be systematically used in various Speech Technology Applications, including Text-to-Speech (TTS) and ASR, the prosodic modelling of Category C or “Prosodically Independent” words is highly sublanguage-dependent and application-specific.

### **References**

- Alexandris, C. 2007. Show and Tell: Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints. *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*. 4552/2007, 13-22.
- Alexandris, C. Fotinea, S-E and Efthimiou, E. 2005. Emphasis as an Extra-Linguistic Marker for Resolving Spatial and Temporal Ambiguities in Machine Translation for a Speech-to-Speech System involving Greek. In *Proc. of the 3rd Intern. Conference on Universal Access in Human-Computer Interaction (UAHCI 2005)*. Las Vegas, Nevada, USA.
- Alexandris, C., Fotinea, S-E. 2004. Discourse Particles: Indicators of Positive and Non-Positive Politeness in the Discourse Structure of Dialog Systems for Modern Greek, *International Journal for Language Data Processing Sprache und Datenverarbeitung*, 1-2/2004, 19-29.
- Hausser, R. 2006. *A Computational Model of Natural Language Communication, Interpretation, Inference and Production in Database Semantics*. Berlin, Springer.
- Moeller, S. 2005. *Quality of Telephone-Based Spoken Dialogue Systems*. New York, Springer.
- Nottas, M., Alexandris, C, Tsopanoglou, A. and Bakamidis, S. 2007. A Hybrid Approach to Dialog Input in the CitizenShield Dialog System for Consumer Complaints. In *Proc. Of Human-Computer Interaction (HCI) 2007*. Beijing, China.
- Schilder, F. and Habel, C. 2001. From Temporal Expressions to Temporal Information. In *Proc. of ACL-2001, Workshop on Temporal and Spatial Information Processing*, 1309-1316, Pittsburgh, Pennsylvania, USA.

## **Saudi accented Arabic voice bank**

Mansour Alghamdi, Fayez Alhargan, Mohamed Alkanhal, Ashraf Alkhairy,  
Munir Eldesouki and Ammar Alenazi

Computer and Electronic Research Institute, King Abdulaziz City for Science and  
Technology

### **Abstract**

The aim of this paper is to present an Arabic speech database that represents Arabic native speakers from all the cities of Saudi Arabia. The database is called the Saudi Accented Arabic Voice Bank (SAAVB). Preparing the prompt sheets, selecting the right speakers and transcribing their speech are some of the challenges that faced the project team. The procedures that met these challenges are highlighted. In the project, 1033 speakers speak in Modern Standard Arabic with a Saudi accent. The SAAVB content was analyzed and the results are illustrated. The content was verified internally by the project team and externally by IBM Cairo and can be used to train speech engines such as automatic speech recognition and speaker verification systems.

Key words: Arabic speech database Saudi

### **Introduction**

Speech databases are essential for training automatic speech recognition systems in addition to other applications such as speaker verification, dialect and language identification. Speech databases are also valuable in linguistic studies especially in phonetics, phonology, typology and sociolinguistics. For these reasons, speech databases of many languages have been collected for many years in many countries: English in Australia, Australian National Database of Spoken Language (ANDOSL) Vonwiller, J.P., et al., (1996); British English speech corpus (WSJCAMO) Robinson et al. (1995); American English, Texas Instrument and Massachusetts Institute of Technology corpus (TIMIT) TIMIT (1990), Macrophone, Bernstein et. al. (1994); Chinese Spontaneous Telephone Speech Corpus on Flight Enquiry and Reservation (CSTSC-Flight), Zheng et al. (2002); Mandarin Across Taiwan (MAT), Tseng et al. (2003); Cantonese, one of the dialects in southern China (Lo et al., 1998); American Spanish (Voice Across Hispanic America) Muthusamy et al. (1995); French, French SpeechDat corpus (FRESCO) Langmann et al. (1996).

Although speech databases have been collected for several languages, Arabic speech databases need more work to cover the dialectal diversity and many Arabic speaking counties remain with almost non-professional speech collections. Saudi Arabia is one of the countries where a speech database that covers its various dialects has not been collected before SAAVB. The

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008,  
25-27 August 2008, Athens, Greece.

area of Saudi Arabia is 1,960,582 sq km. It is located in south west Asia and surrounded from north, east and south by other Arab countries. About 20 million inhabitants live in Saudi Arabia; four fifth of them are native Saudis (Ministry of Economy and Planning, 1999 Census).

A project that aims at collecting speech database faces several obstacles. Finding the right speakers that represent the population is one example. Another example is choosing the linguistic materials that are suitable for both the speaker's culture and useful for training speech recognition systems.

This paper is written to assist the Saudi Accented Arabic Voice Bank (Alghamdi et al., 2003) users and to document the procedures, specifications and contents of SAAVB for those who will be interested in collecting similar speech data in similar environment.

### **Database Design and Recording**

The procedures to collect the data have four phases: 1) designing the Prompt Sheet, 2) selecting the speakers, 3) recording the speech and 4) transcription.

Each speaker is given a 5 page document. The first page has a code that gives each speaker access to the recording lab to record their speech. The code symbolizes the region, city, gender, age, telephone type and calling environment of the speaker. The code is to be used as the name of all SAAVB files, so, the gender, age and other information related to the speaker can be extracted from the name of the files. The second page has the instructions that help the speaker to log into the recording system and complete the required tasks. The remaining 3 pages are the prompt sheets.

A prompt sheet consists of 59 items: 49 read items (83%) and 10 elicited spontaneous responses (17%). All the read items are written in Modern Standard Arabic which is widely used in the media, press and official communication in the Arab world. A unique Prompt Sheet for each speaker; no two Prompt Sheets are the same, is prepared. The only two sentences that appear in every Prompt Sheet are Prompt 41 and 42. These two sentences are designed to record dialectic variations among speakers. The total number of prepared prompt sheets is 1059.

Due to the absence of a dialect map for Saudi Arabia, the research team decides to select a sample from every city in the country. One half of the sample is male, and the other half is female. There are 118 cities in Saudi Arabia spread all over the country area, and the target number of speakers have to be divided among them according to the city population.

### **Results**

The total number of the speakers who participated in SAAVB is 1033 distributed among the following categories:

523 Male (50.63%), 510 Female (49.37%)  
 725 Cellular (70.18%): 252 Quiet environment (34.76%), 252 Noisy environment (34.76%), 221 Moving vehicle (30.48%).  
 308 Fixed telephones (29.82%): 232 Quiet environment (75.32%), 76 s Noisy environment (24.68%),  
 512, 16-30 years old (50.53%),  
 364, 31-45 years old (35.24%),  
 147, 46-60 years old (14.23%).

The total number of words in the transcription files is 302,107 distributed among 60947 text files with an average of 5 words in a file. The dictionary includes 34,961 words. 19,941 words are unique, i. e. they occur only once in the database. The high percentage of the unique words indicates the vocabulary richness of the database where only 8.5% of the transcribed words occur more than 10 times.

The duration of the total recorded speech is 96.37 hours distributed among 60947 audio files (1033 speakers x 59 audio files). This means that the average duration for each speaker is 5.60 minutes and the average duration of each audio file is 5.70 seconds.

## Conclusions

This paper presents a description of the Saudi Accented Arabic Voice Bank (SAAVB); how it is collected and its content. SAAVB has been licensed to IBM to be used to train their speech recognition engine. Currently, it is used by KACST speech team to develop an Arabic speech recognition system.

SAAVB includes 1033 directories with a total of 183,518 files. The total size of SAAVB is 2.59 GByte.

SAAVB is now available as a Saudi Arabic voice bank and can be licensed to be used in research or to develop products when a contract with KACST is signed.

## Acknowledgements

This paper is supported by KACST through SAAVB project number i-e-6-1. The authors would also like to acknowledge the IBM Egypt team, led by Dr. Ossama Emam, for their cooperation with KACST team and for verifying the database.

## References

- Alghamdi, M., F. Alhargan, M. Alkanhal, A. Alkhairi, M. Aldusuqi. 2003. Saudi Accented Arabic Voice Bank. Final Report. Computer and Electronic Research Institute, King Abdulaziz City for Science and Technology.



- Bernstein, J. Taussig, K. And Godfrey, J. 1994. Macrophone: an American English telephone speech corpus for the Polyphone project. *Acoustics, Speech, and Signal Processing*, 1: 1/81-1/84.
- Langmann, D., R. Haeb-Umbach, L. Boves and E. den Os. 1996. FRESCO: The French Telephone Speech Data Collection - Part of the European SpeechDat(M) Project. FRESCO. The Fourth International Conference on Spoken Language Processing. Philadelphia. 1: 1918-1921.
- Lo, W. K., T. Lee and P. C. Ching. 1998. Development of Cantonese spoken language corpora for speech applications. *Proceedings of the First International Symposium on Chinese Spoken Language Processing*. 102-107. Singapore.
- Ministry of Economy and Planning: <http://www.planning.gov.sa/docs/045.htm>
- Muthusamy, Y., E. Holliman, B. Wheatley, J. Picone and J. Godfrey. 1995. Voice across Hispanic America: A telephone speech corpus of American Spanish. *Acoustics, 1995 International Conference on Speech, and Signal Processing, ICASSP-95*. 1: 85-88.
- Robinson, T., J. Fransen, D. Pye, J. Foote and S. Renals. 1995. WSJCAMO: A British English speech corpus for large vocabulary continuous speech recognition. 1995 International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95. 1: 81-84.
- TIMIT: Acoustic-Phonetic Continuous Speech Corpus. DMI. 1990.
- Tseng, C., Y. Cheng, W. Lee and F. Huang. 2003. Collecting Mandarin Speech Databases for Prosody Investigations, The Oriental COCOSDA. Singapore.
- Vonwiller, J. P., et. al., 1996. (Speaker and Material Selection for the Australian National Database of Spoken Language), *Journal of Quantitative Linguistics*, 27.
- Zheng, T. F., P. Yan1, H. Sun, M. Xu, and W. Wu. 2002. Collection of a Chinese Spontaneous Telephone Speech Corpus and Proposal of Robust Rules for Robust Natural Language Parsing. Joint International Conference of SNLP-O-COCOSDA, Hua Hin, Thailand: 60-67.

# **A new Arabic stemming algorithm**

Eiman Tamah AlShammari and Jessica Lin

Department of Computer Science, George Mason University, USA

## **Abstract**

Text processing is a vital step in the information retrieval process, text mining, and natural language processing. It includes several stages, such as normalization, stop word removal, and stemming. Stemming is the process of reducing the lexicon to its root. Due to the different structures and rules in languages, the task of stemming is language-dependent. This research introduces a new stemming algorithm for the Arabic Language. Arabic is one of the most complex languages, both spoken and written. However, it is also one of the most common languages in the world. It is the base from which many other languages are derived. Despite the wide usage of the language, technology and development for Arabic has been limited. The main reason lies within the formulation rules of Arabic, as Arabic language exhibits a very complicated morphological structure. Existing Arabic stemmers suffer from high stemming error-rates. They blindly stem all the words and perform poorly, especially with compound words, proper nouns and foreign Arabized words. The main cause of this problem is the stemmer's lack of knowledge of the word lexical category (i.e. noun, verb, proposition, etc.) This paper presents a new stemming algorithm that relies on Arabic language morphology and Arabic language syntax. The automated addition to the syntactic knowledge reduced both stemming error and stemming cost. Additionally, the new Algorithm automatically creates its own list of proper nouns, and compound words based on the processed corpus.

Key words: Arabic, stemming, morphology

## **Introduction**

The efforts to improve Arabic information search and retrieval compared to other languages are limited and modest, even though the Arabic language is the official language for over 29 countries, in addition to which there are native Arabic speakers scattered all over the world. The barrier to text processing advancements in Arabic is its very complicated morphological structure.

Stemming Arabic documents was done manually prior to TREC (Text Retrieval Conference). The two most effective Arabic stemmers are Larkey's light stemmer (Larkey and Connell 2001; Larkey, Ballesteros and Connell 2002) and Khoja's (Khoja 1999) root-extraction stemmer.

Over-stemming and under-stemming are the main drawbacks of the root-based stemming and the light stemming algorithms, respectively. Over-stemming, under-stemming and mis-stemming are all stemming errors that usually degrade the correctness of stemming algorithms (Baeza-Yates 1992).

To decrease stemming errors, other stemmers add a lookup dictionary to check the roots after stemming. This process is computationally expensive; Al-Fedaghi and Al-Anzi (Al-Fedaghi and Al-Anzi 1989) estimated that there are around 10,000 independent roots and each root word can have prefixes, suffixes, infixes, and regular and irregular tenses.

To mitigate the drawbacks of the previous work on Arabic stemming, we propose an alternative that defines a rule to stem words instead of chopping off the letters. This rule is set by the syntactical structure of the word. For example verbs require aggressive stemming and need to be represented by their roots. Nouns on the contrary only require light suffixes and prefixes elimination.

### Arabic language structure

Arabic language is a semantic language with a composite morphology. The words are categorized as particles, nouns, or verbs. There are 28 letters in Arabic, and the words are formed by linking letters of the alphabet. Letters of the alphabet differ in shape based on their position within the word (i.e. beginning, middle, or end). Unlike most Western languages, Arabic script is written from right to left. Furthermore, proper nouns do not start with capital letters, thus, extracting nouns and proper nouns is a challenging task for machines.

Also, in English, words are formed by attaching prefixes and suffixes to either or both sides of the root. For example the word Untouchables is formed as follows:

Un	touch	able	s
Prefix	Root	First Suffix	Second Suffix

In Arabic, additions to the root can be within the root (not only on the word sides) which is called a pattern. This causes a serious issue in stemming Arabic documents because it is hard to differentiate between root particles and affix letters. For example, for the root “drink” شرب in Arabic, adding the letter “ا” (infix) formed a different words such as شارب “drinker” can be formed by adding the letter “ا” (infixes).

شارب	شرب
ش ا رب	ش ر ب
<b>Drinker</b>	<b>Drink</b>

Figure 1. Arabic infix example.

Suffixes, prefixes and infixes are categorized based on their use. Similar to other Western languages, there are specific suffixes to convert the word from the singular form to the plural form and others to convert from masculine to feminine.

### **Arabic stemming algorithm**

Table 1. Arabic Stemming Algorithm.

---

Input: Arabic document

Output: Stemmed document, Noun Dictionary, Verbs Dictionary

- V: Verb dictionary (one dimensional array sorted alphabetically<sup>1</sup>)
- N: Noun dictionary (one dimensional array sorted alphabetically)
- NSW: Array of stop words proceeding nouns
- VSW: Array of stop words proceeding verbs
- SW: Array of stop words (including both NSW and VSW)

Phase Zero: Remove useless stop words.

Phase One: Simple Noun Identification and Noun Dictionary Generation

Locate words attached to definite articles, and preceded by NSW and flag them as Nouns, Add the identified words to N.

Phase Two: Suffix and Prefix removal

Apply suffix and prefix approach to the entire document. Longest suffixes and prefixes are removed first.

Phase Three: Verbs Identification and Verb Dictionary Generation

Verbs proceeded by VSW are flagged and added to the V

Phase Four: Find all noun tokens

Phase Five: Stop Word Removal : Remove useful and useless stop words

Phase Six : Root Extraction for Verbs

Phase Seven: Roots are extracted by comparing Verbs to Arabic Root patterns; words with missing tags are considered nouns and lightly stemmed.

---

Our novel algorithm consists of different phases. During the first phase, useless stop words are removed to reduce the size of the corpus. Next, we identify nouns by either locating Stop words that always precede nouns (example: the, and a) or words starting by definite articles. These nouns are lightly stemmed by removing suffixes and prefixes and then added to a global nouns dictionary. At this level, these words are flagged as nouns as a preparation to the stemming phase. In parallel to that process we find verbs by locating stop words that always precede verbs. Similar to the nouns, the

verbs are added to a global verb dictionary and tagged as verbs. In Arabic, we cannot have two consecutive verbs, thus any word following a verb is either a stop word or a noun. If the word is not a stop word then the word is added to the noun dictionary and flagged as a noun.

Before we direct a word to the appropriate stemming by the word flag, all the stop words are removed since they offer no further advantage. The document is revisited by categorizing words with missing flags using the noun corpus and the verb corpus as a lookup table. Other words that do not belong in any category will be treated as nouns and stemmed lightly. Table 1 below summarizes the algorithm.

### Notes

<sup>1</sup>For fast lookup, these dictionaries can be implemented using hash tables.

### References

- Al-Fedaghi, S.S. and Al-Anzi, F., 1989. A New Algorithm to Generate Arabic Root-Pattern Forms. *Proceedings of the 11th National Computer Conference and Exhibition*, 391–400.
- Baeza-Yates, R., 1992. Text retrieval: Theory and practice. *12th IFIP World Computer Congress*, 1, 465-476.
- Khoja, S., 1999. *Stemming Arabic Text*. Lancaster, UK, Computing Department, Lancaster University,
- Larkey, L.S., Ballesteros, L. and Connell, M.E., 2002. Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 275-282.
- Larkey, L.S. and Connell, M.E., 2001. Arabic Information Retrieval at UMass in TREC-10. *Proceedings of the Tenth Text REtrieval Conference (TREC-10)*, EM Voorhees and DK Harman ed, 562-570.

# Neurolinguistic aspects of metaphor theory

Georgia Andreou and Ioannis Galantomos

Department of Special Education, University of Thessaly, Greece

## Abstract

The goal of this paper is to examine the neural aspect of metaphor. The neural theory of language stems from the cognitive linguistic research and is an effort to comprehend the way neural circuits affect and shape language and thought. Within this framework it is suggested that metaphor serves as the medium through which cultural, abstract and theoretical concepts are acquired. On the other hand, it is believed that conceptual metaphors are ensembles of neurons located in different parts of the brain and connected by larger neural groups.

Key words: cognitive linguistics, conceptual metaphor, neural circuits, neural mappings

## Introduction

The traditional view holds that metaphor is a figure of speech, a property of words, which is used mainly for rhetorical and artistic purposes (Kövecses 2002). Lakoff and Johnson (1980/2003) in their seminal work “*Metaphors we live by*”, based on a big volume of empirical data, challenged those well established beliefs and pointed at the pervasiveness of metaphor in everyday language. Furthermore, they claimed that the human conceptual system is metaphorical in nature and that abstract thinking relies on metaphor.

In the cognitive linguistic approach, metaphor is a complex phenomenon. In particular, it is suggested that metaphor is conceptual, linguistic, bodily, socio-cultural and neural.

Metaphor is a property of concepts and it refers to the understanding of one less physical/abstract domain (i.e. concept) in terms of a more concrete one. This is called a conceptual metaphor and it consists of a source and a target domain. Thus, in the conceptual metaphor LOVE IS A JOURNEY (conceptual metaphors by convention are written in small capitals), love is the target domain, whereas journey is the source domain. Conceptual metaphors become explicit in language through metaphorical linguistic expressions. An example of the above mentioned conceptual metaphor is the linguistic expression “*we are at a crossroads*”.

Between the source and the target domain there are systematic correspondences which are called mappings. For instance, in the metaphor LOVE IS A JOURNEY the travellers correspond to the lovers. Conceptual metaphors are grounded in embodied experience, that is the way human body operates exerts an influence on the final choice of a certain source



domain to go with a particular target domain. Also, metaphors reflect cultural patterns. In other words, entrenched cultural models affect human understanding of certain concepts.

The neural basis of metaphor is one of the current findings of cognitive researchers. The major assumption is that metaphors are comprehended through the activation of particular neural combinations that are located in the sensorimotor system and in higher cortical areas (Kövecses 2005).

Earlier results, outside the cognitive linguistic research, suggested the contribution of the right hemisphere (RH) to the comprehension of metaphor (e.g. Burgess and Chiarello 1996). Nevertheless, current findings (e.g. Lee and Dapretto 2006) challenge the selective role of the RH and argue against its specialization in metaphor processing.

### **The neural theory of metaphor**

The neural theory of metaphor and by extension of language is a project that is taking place at the University of California at Berkeley aiming at understanding the role of neural circuits in language and thought (Gibbs 2006).

Within this framework, three models were developed. The first one was proposed by Regier (1996) who designed a neural model for acquiring spatial terms in various languages. Regier used topographical maps of the visual field, orientation-sensitive cell systems to compute the orientational aspects of spatial concepts that are related to bodily orientation and center-sensitive receptor fields in order to characterize concepts, such as contact and distance. Regier's model showed very good results in learning these spatial terms.

The second model is Narayanan's (1997) neural theory of metaphor. This project laid an emphasis on metaphorical inferences about events and the verbal category of aspect. Narayanan used computational methods for neural modelling and showed that conceptual metaphors are neuronal combinations located in different parts of the brain and connected by the functions of neural circuits. These combinations are the source and the target domain and the neural circuits are the mappings. Under this perspective, mappings are not anymore simple correspondences between two interrelated domains but are neural mappings that constitute the neural mechanism, which connects the sensorimotor system (where the source domain is located) and higher cortical areas (where the target domain is located).

A last model was developed by Bailey (1998) who attempted to explain the acquisition of hand motion verbs (e.g. push, grasp, wave). Bailey concluded that human body serves as the basis for defining the meanings each hand motion verb has in world's languages. In other words, he

presented proof for the key-role of the sensorimotor system for distinguishing among the semantics of the hand motion verbs.

Among these neural models the most important is the second one, namely Narayanan's neural theory of metaphor which explains adequately the findings of two other advances in cognitive linguistics. The first advance is Grady's (1997) primary metaphor theory which is an attempt to explain the partial nature of metaphorical mappings. For this reason, Grady introduced the notion of primary metaphor in order to address the weakness of the standard metaphor theory to justify why some elements are mapped onto one domain and not others. Primary metaphors create complex metaphor(s) and stand at the higher level of abstraction.

The acquisition of these primary metaphors led Johnson (1997) to conduct an experiment in order to test when children learn the metaphorical sense of the verb *see* as in the sentence "*I see what you mean*". Johnson found that the literal meaning of this verb was acquired first. Then, a conflation period followed where both literal and metaphorical meanings were present and active and finally pure metaphorical cases were learned. Johnson's findings serve as the second advance.

The results of both Grady's and Johnson's experiments are best explained by Narayanan's neural theory of metaphor. The primary metaphors are learned on the basis of neural mappings that connect the source and the target domain and are active during the conflation period. New metaphors are learned through the establishment of new neural mappings and not the emergence or the creation of a copy machinery. One implication is that there is no need of overrides because metaphors are acquired when two experiences occur at once. If a neural mapping in the target domain leads to a contradiction then it will not be learned and it will be inhibited.

According to Lakoff and Johnson (1980/2003) metaphor in the neural theory of language has certain advantages. First, it explains the way primary metaphors are learned. Second, it explains the ubiquity of metaphors. Third, there is no need of overrides. Fourth, metaphors fit well with the other aspects of this theory. Fifth, it can explain the dynamic role and the use of metaphor in everyday language and finally it offers a computational model for the study of metaphor in discourse.

## Conclusions

In sum, metaphor, within the neural theory of language, becomes a neural mechanism. An entailment of this view is that the most abstract products of the human conceptual system, namely the primary metaphors, are acquired automatically and unconsciously during the period of conflation. People have no choice in avoiding this because of the stable and constant way neural mappings operate and connect sensorimotor system and higher

cortical areas of the brain each time they attempt to understand abstract concepts in metaphorical terms.

## References

- Bailey, D. 1998. *Getting a Grip: A Computational Model of the Acquisition of Verb Semantics for Hand Actions*. Unpublished doctoral dissertation. University of California, Berkeley, USA.
- Burgess, C. and Chiarello, C. 1996. Neurocognitive Mechanisms Underlying Metaphor Comprehension and Other Figurative Language. *Metaphor and Symbolic Activity* 11, 67-84.
- Gibbs, R.W. Jr. 2006. *Embodiment and Cognitive Science*. Cambridge: CUP.
- Grady, J. 1997. *Foundations of Meaning: Primary meanings and primary senses*. Unpublished doctoral dissertation. University of California, Berkeley, USA.
- Johnson, C. 1997. Metaphor vs conflation in the acquisition of polysemy. The case of SEE. In Hiraga, M.K., Sinha, C. and Wilcox, S. (eds.), *Cultural, typological and psychological issues in cognitive linguistics*. *Current Issues in Linguistic Theory*, vol. 152, 155-169. Amsterdam, John Benjamins.
- Kövecses, Z. 2002. *Metaphor. A Practical Introduction*. Oxford: OUP.
- Kövecses, Z. 2005. *Metaphor in Culture. Universality and Variation*. Cambridge: CUP.
- Lakoff, G. and Johnson, M. 1980/2003. *Metaphors we live by*. Chicago and London: The University of Chicago Press.
- Lee, S.L. and Dapretto, M. 2006. Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *NeuroImage* 29, 536-544.
- Narayanan, S. 1997. *Moving Right Along: A Computational Model of Metaphoric Reasoning about Events*. Unpublished doctoral dissertation. University of California, Berkeley, USA.
- Regier, T. 1996. *The Human Semantic Potential*. Chicago: The University of Chicago Press.

# **Prosodic variation in L2: a case of Germans speaking English**

Volha Anufryk<sup>1</sup>, Matthias Jilka<sup>2</sup> and Grzegorz Dogil<sup>1</sup>

<sup>1</sup>Institute for Natural Language Processing, Stuttgart University, Germany

<sup>2</sup>Institute for English Linguistics, Stuttgart University, Germany

## **Abstract**

The present study investigates prosodic variation as realized by L2 speakers of varying pronunciation ability in comparison with native speakers of English. The results demonstrate the distribution of the rising contours on both the phonological (ToBI frequencies) and phonetic (values for six parameters of the F0 curve) levels. The rising contours and pitch accents have a wider distribution in German productions, and are therefore closer to the German prosodic pattern, as opposed to the native realizations. Another peculiarity concerns the F0 peak frequency parameter in L\*H accents as realized by the below-average informants. Their values are significantly different from those of the native and average speakers. Further phonetic differences are to be tested for consistency effects on the subsequent research stages.

Key words: prosodic variation, second language acquisition, language ability, F0 contour

## **Introduction**

Most modern theories of language acquisition suggest that L2 is a reduced system characterized by a basic variety (Klein and Perdue 1997) of language means, i.e. a limited set of exemplars (e.g. Lacerda 1995) which a speaker reproduces, once he or she has been exposed to them.

Thus, it was claimed, for example, that Finns vary within a narrower pitch range in Russian (Ullakonoja 2007) and Americans in Mandarin (Bent 2005) than the respective native speakers. A number of other investigations determined an opposite trend on this matter: more variation was found in the tonal structures of Americans speaking Japanese (Ueyama 2000), as well as in the vowel representation of Spanish speakers of English (Wade et al. 2007).

The absence of consensus on the issue of variation across L1 and L2 implies that the outcome largely depends on the languages and phenomena involved.

It would be therefore interesting to have a closer look at the prosodic subsystem, as it per se allows of a considerable degree of variation within its categories. Irrespective of the linguistic component, variation seems to have a relation to an individual's language ability, as it requires an expansion of

the basic variety, on the one hand, and accommodation to the L2 variation pattern, on the other.

A cross-linguistic study was conducted to test the above hypotheses, whereby a detailed analysis of F0 variation was carried out.

## Method

Data consisted of read speech samples produced by 30 native German speakers whose pronunciation ability had been initially defined as excellent, average and below average (12, 10 and 8 speakers in each group, respectively), based on the tests performed as part of the DFG funded 'Language Talent and Brain Activity' Project. Corresponding recordings by 12 native English speakers were available for comparison. The whole corpus was segmented into syllables, followed by an automatic extraction of F0 values for each syllable. The respective F0 values were parametrized using the PaIntE method (Möhler 2001), which describes the F0 curve in terms of five basic parameters: steepness of the rising and falling sigmoids; the alignment of the function within a syllable; amplitudes of the rising and falling sigmoids and the frequency of the F0 peak.

Finally, intonation events were labeled manually in accordance with the ToBI convention.

## Results

The initial extraction of global PaIntE parameters, i.e. without regard to the ToBI accents, yielded a clear peculiarity in the realization of the rising sigmoid by below-average German speakers, as compared to all the other subjects. This result was taken as an incentive to analyze the rising F0 contour in more detail.

Clear differences between the groups were evident already from descriptive statistical analysis. The rising F0 contour (high boundary) had a much wider distribution in German realizations than it was employed by the native speakers. The frequency was highest for the average group, followed by below-average speakers. Informants of excellent ability approached native-like performance, but their values were still much higher.

Given the fact that the speech samples were taken from a neutral text for reading, the German speakers' preference for a rising contour can be interpreted as a typical pattern in this type of speech.

Another peculiarity concerned the distribution of the 'rising' pitch accents L\*H and LH\* in pre-boundary position. These ToBI events were also more typical of German samples, with the highest percentages for the average and below-average groups and the lowest for the native speakers.

Table 1 demonstrates all of the above findings in percentage values.

Table 1. Distribution of rising contours and pitch accents.

Group	High boundary	L*H	LH*
below-average	37,6	16,7	1,7
average	48,8	17,4	1,5
excellent	31,3	10,4	3,0
native	10,2	6,1	2,6

As a next step we looked at the separate PaIntE parameters of L\*H and LH\* pitch accents to see if there were phonetic differences in their realization. The values for each parameter were normalized groupwise and substituted for the corresponding z-scores. Then we compared the samples by means of the Kolmogorov-Smirnov test. Whereas no significant differences were found for the LH\* accent, the p-values for each two samples, i.e. speaker groups, of L\*H events lay below 0.05 for at least a few parameters.

It is notable in this respect that the frequency of F0 peak was only significantly different for the below-average group opposed to the native and average speakers. This finding is also represented in the following boxplot: the values of the below-average informants stand out as most centred and scattered at the same time.

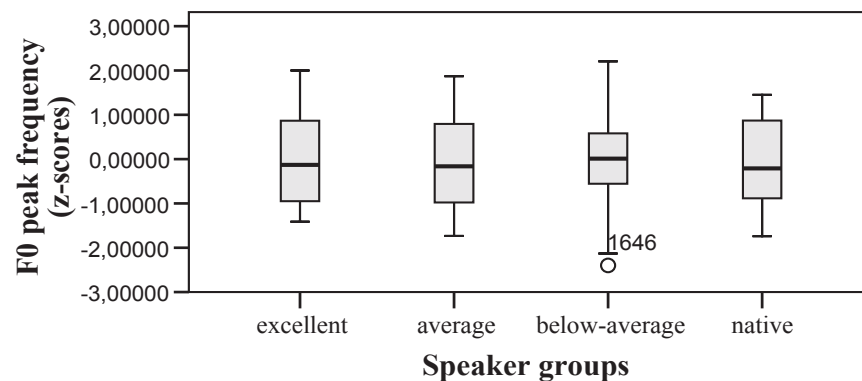


Figure 1. F0 peak frequency of L\*H accent.

### Discussion

The results of the present study confirm the initial hypothesis of a correlation between prosodic variation and language ability. Observable differences were found both in the distribution of ToBI categories and in the realization of individual F0 curve parameters. While phonological findings were expected and show a clear trend towards the native patterns in German productions (a wider distribution of the rising contours was described earlier



by Anderson (1979)), elaborate phonetic peculiarities of PaIntE parameters across the speaker groups require further exploration for consistency effects.

### **Acknowledgements**

The data for the current study was taken from the corpus of the “Language Talent and Brain Activity” Project supported by the German Research Foundation.

### **References**

- Anderson, K. O. 1979. On the contrastive phonetics of English and German intonation. *Festschrift für Otto von Essen anlässlich seines 80. Geburtstages*. In H-H. Wängler (ed.), *Hamburger Phonetische Beiträge* 25, 25-35
- Bent T. 2005. Perception and Production of Non-Native Prosodic Categories. Doctoral Dissertation, Northwestern University
- Klein W. and Perdue C. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?), *Second Language Research* 13,4; pp. 301 - 347
- Lacerda F. 1995. The Perceptual-magnet effect: an emergent consequence of exemplar-based phonetic memory. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm*, vol.2, pp. 140-147
- Möhler G. 2001. Improvements of the PaIntE model for F0 parametrization. *Research Papers from the Phonetics Lab, AIMS Universität Stuttgart*.
- Möhler G. and Mayer J. 2001. A Discourse model for pitch-range control. 4th ISCA Workshop on Speech Synthesis Perthshire, Scotland.
- Ueyama M. 2000. Prosodic Transfer: An Acoustic Study of L2 Japanese & L2 English. Doctoral Dissertation, UCLA.
- Ullakonoja R. 2007. Comparison of pitch range in Finnish (L1) and in Russian (L2). *Proceedings of the XVIth Congress of Phonetic Sciences, Saarbrücken*, pp. 1701-1704
- Wade T., Jongman A. and Sereno J. 2007. Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, vol. 64, pp. 122-144.

# The influence of top-down expectations on the perception of syllable prominence

Denis Arnold and Petra Wagner

Institute of Communication Sciences, University of Bonn, Germany

## Abstract

In our study we use the experimental framework of *priming* to manipulate our subjects' expectations of syllable prominence in sentences with a well-defined syntactic and phonological structure. It shows that it is possible to prime prominence patterns and that priming leads to significant differences in the judgment of syllable prominence.

Key words: top-down, priming, syllable prominence, perception

## Introduction

Experimental phonetics has long sought to define accurate acoustic correlates of syllable prominence. Findings of several studies e.g. Eriksson (2001), Wagner (2005), indicate that syllable prominence ratings can be affected by top-down processes in addition to acoustic cues. However, a systematic investigation such influences has not been carried out yet. In our study we use the experimental framework of *priming* to manipulate our subjects' expectations of syllable prominence in sentences with a well-defined syntactic and phonological structure. We examine if priming leads to different ratings of syllable prominence thus gaining better insight into the role that top-down expectations play for the perception of syllable prominence.

We describe two experiments. The first experiment uses an intraindividual design. Due to some problems, we carried out a second experiment with a interindividual design with four groups.

## Experiment 1

32 subjects were asked to rate the syllable prominence of 44 sentences presented via headphones with the help of ten sliders on a computer screen (cf. Fig. 1). The slider had to be moved to the top of the scale, if the syllable was rated maximally prominent. In case of a completely non-prominent syllable, the slider had to be kept in the lowest position. The subjects were encouraged to use the full range of the sliders.

In a training phase, the subjects were familiarized with the experimental setting. Then, six test-sentences were rated, followed by a distraction task. In the following *priming* stage we presented 24 sentences with equal

syntactic and similar semantic structure for each of the initial test-stimuli. All priming sentences belonging to one test sentence shared the prosodic pattern. However, this differed from the pattern of the pertinent test sentence in the accentuation of one particular syllable. The test-sentences were presented again in the last test stage. If the priming is successful a significant difference between the first and second rating of the test sentences should be the result.

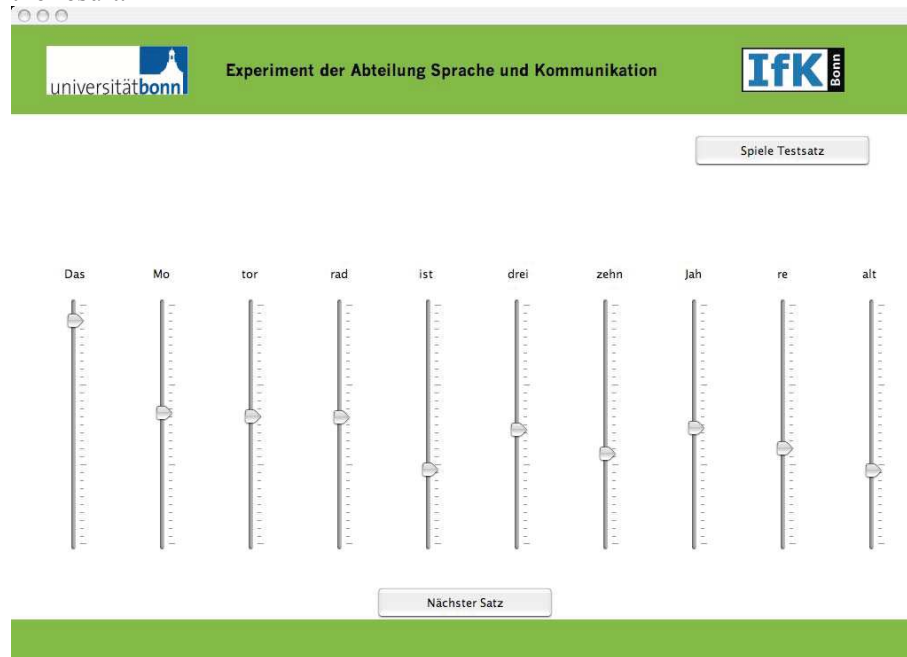


Figure 1. The GUI for the prominence rating.

One finding is that the average ratings of test sentences were much lower in the second rating. (cf. Figure 2 for an example)

Table 1. Results of Experiment 1.

Condition A	Condition B
Sentence 1 $t(31) = 2.11, p < .05$	Sentence 1 $t(31) = 0.9, p = .8125$
Sentence 2 $t(31) = 2.0271, p < .05$	Sentence 2 $t(31) = 1.6515, p = .9456$
Sentence 3 $t(31) = 1.9823, p < .05$	Sentence 3 $t(31) = 2.1573, p = .9806$

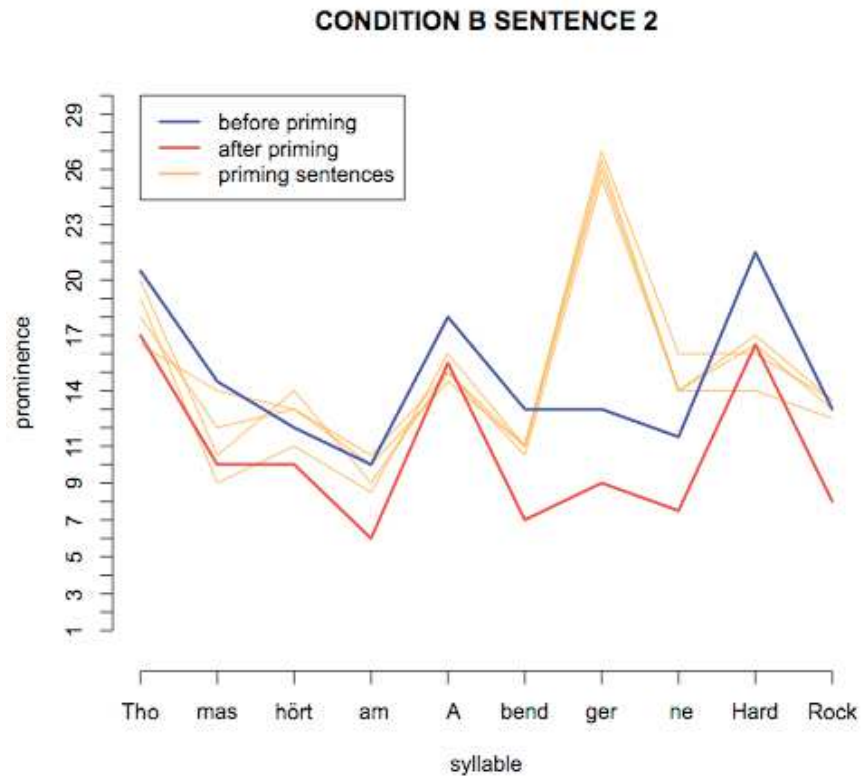


Figure 2. Prominence rating. A test sentence in condition B.

The manipulated syllables show the predicted difference. The results look promising, if one looks only at the manipulated syllable. When looking at the differences between the manipulated sentences and their neighbors, no significant difference is found. This lead to the second experiment where we used a four group design. This should help to avoid a repetition of the presentation of the test sentence and make the duration of the experiment much shorter.

## Experiment 2

For this experiment 72 subjects where asked to rate the syllable prominence of 20 sentences. The same interface was used for presentation and rating as in the first experiment.

There were two conditions with two groups. Each group was primed with a different set of priming material and exposed to the same test sentence in the end of the test. The test sentences where compared.

We mostly found significant differences in the ratings in both conditions. (cf. Table 2) In the group, where the priming material contained one stressed syllable, we found that the ratings of the not manipulated syllables were lower than in the other group for condition A.

Table 2. Results of Experiment 2.

Condition A	Condition B
Sentence 1 $t(33.65) = -3.5608, p < .01$	Sentence 1 $t(33.529) = 2.0652, p < .05$
Sentence 2 $t(27.353) = -2.1909, p < .05$	Sentence 2 $t(31.096) = -0.0365, p = .5144$
Sentence 3 $t(28.297) = -1.6834, p = .05165$	Sentence 3 $t(31,737) = 2.156, p < .05$
Sentence 4 $t(24.103) = -1.8616, p < .05$	Sentence 4 $t(32.835) = 0,7846, p = .2192$

### Conclusion and Outlook

We were able to show, that the priming paradigm is well suitable for the research of top-down expectations. The results of this study give further support to the hypothesis, that top-down expectations have an impact on the rating of syllable prominence.

Further studies will look how different words and positions alter the effect size. Another goal is the estimation of the amount of influence of the top-down expectation on the rating of syllable prominence.

### References

- Eriksson, A., Grabe E. And Traunmüller, H. 2002. Perception of syllable prominence by listeners with and without competence in the tested language. *Proceedings Speech Prosody 2002, Aix-en-Provence*, 275-278.
- Wagner, P. 2005. Great expectations - introspective vs. perceptual prominence ratings and their acoustic correlates. *Proceedings of INTERSPEECH 2005, Lisbon*, 2381-2384.

# **Overspecification in action-oriented discourse: task importance affects the production of overspecifications and overspecifications increase identification efficiency in perception**

Anja Arts<sup>1</sup>, Alfons Maes<sup>1</sup>, Leo Noordman<sup>1</sup> and Carel Jansen<sup>2</sup>

<sup>1</sup>Faculty of Humanities, Tilburg University, the Netherlands

<sup>2</sup>Faculty of Arts, Radboud University, the Netherlands

## **Abstract**

This paper discusses the effect of overspecified reference as a communicative strategy in action-oriented discourse. In a perception and a production experiment, based on identical instructive environments, participants were asked to identify an object after reading a minimally specified or overspecified expression, or to produce a referential expression in a low-importance or a high-importance context. Overspecification shows to be a pervasive instructive production strategy, which mostly affects the addressee's identification task in a positive way.

Key words: overspecification, instruction, coherence, reference

## **Introduction**

Cognitively oriented research of discourse reference aims to explain and predict the form and specification level of referential expressions in discourse. It centers on the discourse structural factors that affect the cognitive status - and consequently also the specification level - of discourse referents (Ariel, 1991; Gundel et al., 1993).

The mechanism of referential coherence is complicated considerably in less 'autonomous' conditions of language use, in particular when language is more closely integrated with perception, action and participants' interaction. We address the issue of overspecification (i.e. providing more referential information than is necessary for unambiguous identification) as a referential strategy in task and action oriented discourse.

Instructors in a non-feedback situation tend to display a considerable degree of referential overspecification in explaining procedures and actions to be executed (Maes et al., 2004). Clark and Wilkes-Gibbs (1986) offer an elegant and plausible explanation of this overspecification strategy in formulating the *principle of distant responsibility*<sup>1</sup>. This producer-oriented explanation, however, leaves at least two questions unanswered. First, how does this producer's strategy relate to the referential needs of the addressee? Is it to be considered a useless but harmless addition, or is it beneficial and really required for the current purposes of the exchange? Second, what are

the situational and interaction-oriented parameters that determine the nature and the extent of the overspecification?

## Method

The research questions were addressed in two experiments, in an identical instructive context (for full details see Arts, 2004). For this purpose, computer screens were designed, consisting of a panel with four objects. The objects could differ in location, shape, color and size. Figure 1 provides an example computer screen.

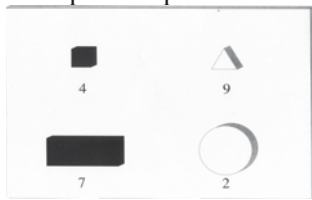


Figure 1. Example computer screen.

## Material

*Perception experiment* - The design of the computer screens accommodated the use of seven minimally specified expressions (Table 1). These expressions contained just the necessary information for identification of the referent, but not more information. Every minimally specified expression could be expanded using one or more of the remaining types of information units still available for reference. This led to a total of twenty possible overspecified expressions.

Table 1. Minimally specified expressions and information units.

ident. nr.	shape	size	color	vert. axis	horiz. axis	example
1	√					the round button
2		√	√			the large white button
3		√		√		the large button at the top
4		√			√	the large button on the left
5			√	√		the white button at the top
6			√		√	the white button on the left
7				√	√	the button at the top left

*Production experiment* - For the production experiment the computer screens were adapted in two ways. Firstly, the numbers that identified the objects were removed, and secondly, one of the objects was marked with an X.



## Procedure

*Perception experiment* – 56 students were presented with the computer screens and asked to identify one of the four objects on the basis of one of the twenty-seven possible types of referential expressions (seven minimally specified and twenty overspecified).

*Production experiment* - 25 students in the low-importance condition and 27 students in the high-importance condition were presented with the computer screens and asked to refer to one of the four objects (high-importance: “indicate which button the surgeon has to push next”; low-importance: “type in which object is marked on the screen”).

## Results

*Perception experiment* - The identification time was measured in milliseconds. The data were analyzed using one-way analyses of variance with level of specification (*minimally specified*, *overspecified*) as within-subjects factor. The results showed that an expansion of a minimally specified expression with a reference to the vertical axis or a reference to both axes led to shorter identification times. An expansion with a sole reference to the horizontal axis did not affect the identification time. Furthermore, an expansion of a minimally specified expression with additional size, color, or shape information only led to shorter identification times if the expansion rendered the referential expression exhaustive in attribute-based information: when shape as well as color as well as size were mentioned.

*Production experiment* - The data were analyzed using a t-test for independent samples. In analyzing the referential expressions that were produced, the focus was on the total number of information units used (Figure 2) and on the type of information units used (Figure 3 and Figure 4) in building the referential expression.

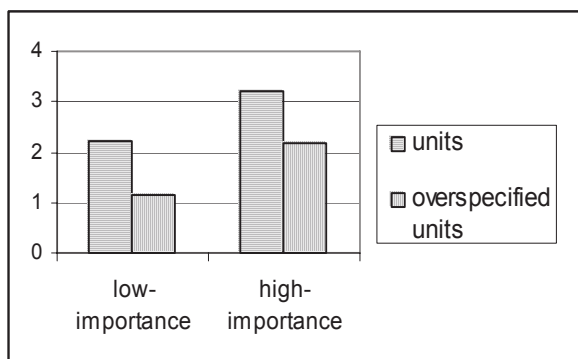


Figure 2. Total units and overspecified units per description.

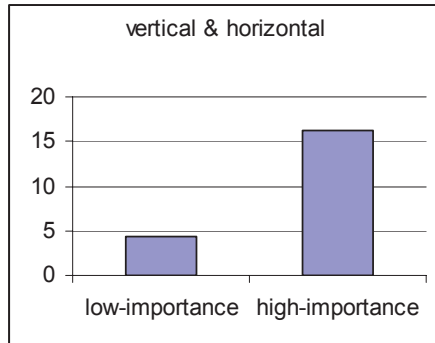


Figure 3. Objects (max. 30) described by referring to both location axes.

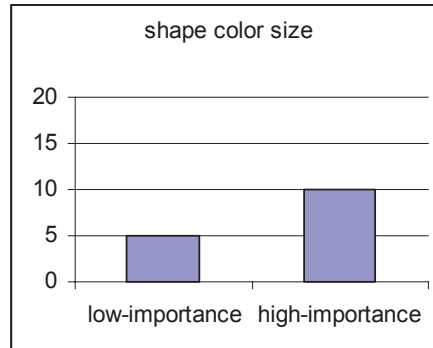


Figure 4. Objects (max. 30) described by referring to shape, color and size.

### Discussion

The processing task (perception) is affected by the use of minimally specified versus overspecified referential expressions, and referential production is affected by the producer's task orientation. Though overspecification undoubtedly increases reading time, it does not decrease the identification efficiency. These results indicate that the producer's task orientation and distant responsibility, rather than possibly the complexity of an instructive task or the assumed experience level of an addressee, are responsible for the overspecification strategy.

### Notes

- <sup>1</sup> The *principle of distant responsibility* refers to the fact that, in a non-feedback situation, the language recipient can not ask for clarification. This may cause the language producer to be highly informative (Clark and Wilkes-Gibbs, 1986).

### References

- Ariel, M. 1991. The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16, 443-463.
- Arts, A. 2004. Overspecification in instructive texts. Doctoral dissertation, Tilburg University, the Netherlands.
- Clark, H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22, 1-39.
- Gundel, J., Hedberg, N. and Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.
- Maes, A., Arts, A. and Noordman, L. 2004. Reference management in instructive discourse. *Discourse Processes* 37, 117-144.

# Prosodic phrasing in German sentence production: optimal length vs. argument structure

Petra Augurzky

Institute for Natural Language Processing, University of Stuttgart, Germany

## Abstract

In the last few decades, language research has been characterized by an increasing interest in the role of prosody in sentence processing. Besides signalling disambiguation, prosodic phrasing has been shown to be sensitive to balancing demands. The present experiment examined the interaction of both domains in the production of *DPI-DP2-V* constructions in German. Acoustic analyses indicate that the need to signal argument structure is mediated by prosodic balance: A general tendency of grouping arguments together with their respective heads was observed, which was increased in saliency when prosodic balance supported such a phrasing.

Key words: sentence production, argument structure, prosodic balance

## Introduction

In the last few decades, language research has been characterized by an increasing interest in the role of prosody in sentence processing. For example, it has been examined experimentally whether prosodic phrasing is used as a disambiguating cue in sentence comprehension and production (e.g. Schafer et al., 2000; Snedeker and Trueswell, 2003). Besides signalling disambiguation, prosodic phrasing has been shown to be sensitive to prosodic balance (i.e. a tendency for realizing prosodic phrases of comparable length, see Fodor, 2002).

Up to now, research on the relation between balance and ambiguity resolution has predominately focused on adjunct attachment. For example, it has been demonstrated that the interpretation of so-called “relative clause attachment ambiguities” is sensitive toward the length of the potential attachment sites, as well as the length of the relative clause (e.g. Fernández, 2003). However, constraints on prosodic phrasing might be more closely related to the processing of sentential core elements (e.g. Schafer et al., 2000). Moreover, considering current phonological theory, length has often been rather loosely defined as an experimental factor. One notable exception comes from a study in Hebrew, in which the exact number of accented elements and their impact on balance was controlled (see Shaked, 2007). Given this background, the present experiment was intended to reveal the mechanisms that constrain the production of argument structure ambiguities in German.

### Sentence production study

The present study explored the interaction between balance and argument structure in German sentence production ( $n = 10$  German speakers). For this purpose, the acoustic properties of *DPI-DP2-V* constructions were analyzed.

### Materials and method

Materials were tightly controlled considering the number of accented elements. Depending on the transitivity information of the clause-final verb, experimental sentences could be interpreted as involving two arguments (TE and TL) or as a possessive construction (IE and IL).

Table 1. Experimental stimuli. Conditions are coded with respect to TRANSITIVITY (first letter: T = transitive; I = intransitive) and BALANCE (second letter: E = expected early break; L = expected late break).

Condition	Example
TE	..., dass g�estern der F�hrer # der R�chterin gedr�ht hat. ..., that yesterday the driver the judge threatened <sub>TRANS</sub> has ..., “that the driver threatened the judge yesterday”
TL	..., dass der F�hrer der R�chterin # zum wiederh�lten Mal gedr�ht hat. ..., that the driver the judge to-the repeated time threatened <sub>TRANS</sub> has ..., “that the driver threatened the judge repeatedly”
IE	..., dass g�estern der F�hrer # der R�chterin geg�lft hat. ..., that yesterday the driver the judge golfed <sub>INTRANS</sub> has ..., “that the driver of the judge golfed yesterday”
IL	..., dass der F�hrer der R�chterin # zum wiederh�lten Mal geg�lft hat. ..., that the driver the judge to-the repeated time golfed <sub>INTRANS</sub> has ..., “that the driver of the judge golfed repeatedly”

Based upon accentual information, conditions TE and IE should preferably exhibit a prosodic boundary between the two DPs (cf. Selkirk, 2000), as indicated by the '#' diacritic. By contrast, the accent structure in TL and IL should lead to a phrasing where a prosodic boundary follows the second DP. Whereas a break separating both DPs is associated with a one-argument reading, a break after the second DP is associated with a possessive interpretation (e.g. Augurzky and Schlesewsky, under revision), thus leading to a potential tension between length and argument structure.

### Results

Acoustic analyses (F0 and duration) for the single items in the complement clause are given in Table 1. Generally, consistent acoustic differences between conditions were restricted to the boundary regions. For N1 duration, a main effect of TRANSITIVITY was found ( $F(1,9) = 6.64$ ;  $p < .04$ ). For N2

duration, a main effect of TRANSITIVITY was found ( $F(1,9) = 5.04$ ;  $p = .05$ ), as well as an effect of BALANCE ( $F(1,9) = 11.7$ ;  $p < .01$ ). Differences in maximal F0 values were restricted to N2: An effect of BALANCE was found ( $F(1,9) = 66.84$ ;  $p < .001$ ).

Table 2. Mean durational values and F0 maxima for the single words in each condition.  $W_{acc}$  stands for the final accented word in the complement clause.

Words	Mean Duration in ms				F0 maxima in Hz			
	TE	TL	IE	IL	TE	TL	IE	IL
dass	201	197	201	196	121	118	119	123
D1	127	137	124	133	130	136	130	136
N1	510	508	440	428	155	160	154	158
D2	125	123	114	112	129	131	132	133
N2	499	552	536	588	143	156	144	159
$W_{acc}$	507	506	508	510	190	197	196	207

An additional analysis was carried out in order to investigate F0 contours of N1 and N2. No significant effects were observed for N1. F0 contours for N2 are given in Figure 2. Conditions differ from point 4 on (point 4 to point 8: All F values  $> 9$ ).

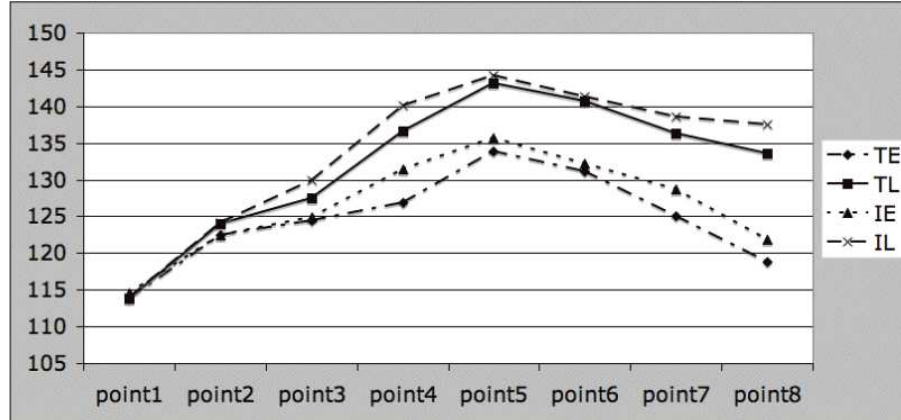


Figure 1. F0 contour for the second noun in the complement clause. For this analysis, the noun was divided into seven fragments of equal length. Mean F0 values at each resulting measurement point were calculated.

## Discussion

The present experiment examined whether prosodic balance and the need to signal argument structure constrain the production of German *DP-DP-V*

ambiguities. As durational analyses indicate, sentences were disambiguated with respect to their argument structure: N1 duration was increased in the transitive conditions, whereas N2 duration was increased for the intransitive conditions. Thus, generally, arguments seem to be planned together with their heads (cf. Watson and Gibson, 2004). Interestingly, acoustic differences due to argument structure were restricted to durational parameters. No F0 differences between transitive and intransitive constructions were observed. This finding indicates that prosodic phrase boundaries due to argument structure are not reflected by tonal properties. In addition, prosodic balance also affected phrasing. However, this effect only occurred at the second potential boundary region, i.e., on N2. When the ideally balanced output biased toward a late break, N2 duration was significantly increased. Moreover, an F0 increase was observed for the late boundary condition. Finally, whereas durational differences were more salient for the transitivity effect (75 ms on N1 and 73 ms on N2 vs. 53 ms on N2 for the balance effect), balance was additionally realized by the F0 parameter. Whether these differences can be interpreted in terms of reflecting different phonological levels (e.g. major vs. minor phrases) has to be examined in future phonological analyses of the materials, as well as in additional perception studies, which are currently underway.

In sum, we found evidence for a prosodic disambiguation of argument structure ambiguities which is additionally constrained by prosodic balance.

## References

- Augurzyk, P. and Schlesewsky, M. (under revision). Prosodic phrasing and transitivity in head-final sentence comprehension – ERP evidence from German ambiguous DPs.
- Fernández, E.M. 2003. Bilingual sentence processing: Relative clause attachment in English and Spanish. Amsterdam: John Benjamins Publishers.
- Fodor, J.D. 2002. Psycholinguistics cannot escape prosody. *Proceedings of Speech Prosody 2002*, 83-90, Aix-en-Provence, France, April 11-13.
- Schafer, A.J.; Speer, S.R.; Warren, P. and White, S.D. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29, 169-182.
- Selkirk, E. 2000. The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 231-261). Dordrecht: Kluwer.
- Snedeker, J. and Trueswell, J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48, 103-130.
- Shaked, A. 2007. Competing syntactic and phonological constraints in Hebrew prosodic phrasing. *The Linguistic Review*, 24, 169-199.
- Watson, D. and Gibson, E. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19, 713-755.

# Coarticulation in non-native speakers of English: /əIv/-sequences in non-proficient vs. proficient learners

Henrike Baumotte and Grzegorz Dogil

Experimental Phonetics Group, Institute for Natural Language Processing,  
Universität Stuttgart, Germany

## Abstract

This is an acoustic study of the production of English /l/ by Standard German speakers. Previous research categorized these speakers as either non-proficient or proficient with regard to their phonetic abilities. Coarticulation differences might be one of several reasons for less proficient speakers not to be able to overcome their foreign accent. The pattern of consonantal velarization influences the degree of /əIv/-coarticulation within the group of L2 speakers. Significant  $F_2$  and  $F_v$  values for non-proficient vs. proficient speakers suggest that proficiency can be characterized due to English stimuli being articulated with more active tongue dorsum control (more velarization) which does not appear as such in German language.

Key words: coarticulation, coarticulatory resistance, foreign accent, acoustic measurements, formant values

## Introduction

In running speech, articulatory gestures overlap in time, leading to interaction between successive phonetic segments, referred to as coarticulation. Previously, several studies investigated coarticulation and found cross-language differences (Öhman 1966; Manuel 1990; Recasens, Fontdevila and Pallarès 1995). Recasens, Fontdevila and Pallarès (1995) reported for German the value for  $F_2$  in /l/ to be lower overall than in other languages. The tongue dorsum is more constrained for German non-velarized [l] and thus less sensitive to coarticulatory effects from, e.g., /i/ or /a/. The authors compared German with Catalan production and observed greater dorsal contact at the palatal zone for German [l] than for Catalan [l]. In line with the surrounding formant frequencies for the vowel /a/, consonantal effects on  $F_2$  for /ə/ are also large because no defined vocal-tract shape is necessary for the production of /ə/; this is why schwa is highly sensitive to coarticulation (Recasens 1985).

## Methods

Formant frequency data were collected for the sequence /dZəI[a, i, y:, u]t/ spoken in as native-like as possible Standard Southern British English, embedded in carrier sentences (I have said [...] twice.) with stress on the

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.



second syllable. This speech material was read five times by each of 21 native German speakers resulting in 420 tokens (1 consonant x 4 vowels x 5 repetitions x 21 speakers). Subjects took part in extensive tests of phonetic language ability based on the large scale DFG project “Language Talent and Brain Activity”, which assessed pronunciation talent in English before. In the beginning of each session subjects were instructed to repeat a small text presented by a native speaker, to help speakers to switch from one language mode to the other. Digital recordings were made at a 16 kHz sampling rate in a sound-attenuated room in the phonetics laboratory at the Universität Stuttgart. The data were then segmented at the phon level by automatic forced alignment (Aligner, St. Rapp, IMS) and formant frequencies were measured every 10 ms with the ESPS *formant* program.  $F_2$  and  $F_1$  were extracted from the middle of the steady state in /ə/.

### Results and discussion

Based on previous studies we predicted that velarized native-like English [ɫ] should show less coarticulatory effects on /ə/ before /l[a, i, y:, u]/ than non-velarized less proficient articulated English [l] (no active tongue dorsum gesture).  $F_2$  frequency and the frequency distance between  $F_2$  and  $F_1$ ,  $F_v$ , which considers also the contribution of  $F_1$  known to be inversely related to velarization (Recasens, Fontdevila and Pallarès 1995: 41), served as indicators for the degree of consonantal velarization. Following these assumptions  $F_2$  and  $F_v$  in less proficient English speakers should be lower than in proficient speakers.

Statistical analysis showed significant  $F_2$  and  $F_v$  differences between non-proficient and proficient non-native English speakers; while for proficient speakers mean  $F_2$  values (*mean value*: 1874,70 Hz) were higher than for non-proficient speakers (*mean value*: 1776,68 Hz).

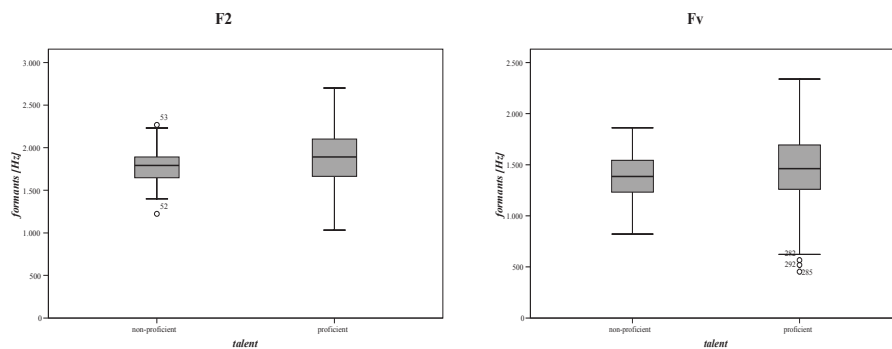


Fig. 1. Distribution of  $F_2$  (left) as well as  $F_v$  in non-proficient vs. proficient speakers (right).

Table 1. *t*-test-results after comparison of  $F_2$  as well as  $F_v$  in non-proficient vs. proficient non-native English speech.

comparison	T =	significance
$F_2$	-2,354	$p = 0,019$
$F_v$	-2,084	$p = 0,037$

In non-proficient speakers  $F_v$  was lower (*mean value*: 1371,57 Hz) than in proficient speakers (*mean value*: 1464,83 Hz). Oh (2008) investigated coarticulation differences in native vs. non-native French as well as English speakers and concluded that more experienced learners developed more native-like degrees of coarticulation than less experienced learners did. In accordance to Oh's results, these data suggest that proficient speakers acquire better the fine-grained language-specific patterns of coarticulation. Subjects categorized as less proficient might not automatically be able to enlarge their stored phonetic features after having heard a sound which is not similar to those existing in their mother tongue. As a consequence, during L2 production not as many exemplars as in proficient speakers can be activated (Pierrehumbert 2001). The use of tongue dorsum control, tongue dorsum fronting and raising might correlate with perception abilities. Keating (1990) proposed that language-specific phonetic details of each language, coarticulation and its amount are specified separately in the grammar of each language. Therefore, amount of /l/-velarization might not necessarily be exploitable for language learners. In future work, we would like to further unravel whether coarticulation differences in non-proficient vs. proficient speakers occur due to perceptual distinctiveness constraints or to independent learning of coarticulatory patterns.

### Acknowledgements

This work was supported by the German Science Foundation, Graduate School 609, Universität Stuttgart, Germany. We would like to thank PD Dr. Wolfgang Wokurek for support and advice.

### References

- Bladon, R.A.W. and Al-Bamerni, A. 1976. Coarticulation resistance in English /l/. *Journal of Phonetics*, 4: 137-150.
- Keating, P.A. 1990. The window model of coarticulation: Articulatory evidence. In: Kingston, J. and Beckman, M. E. (eds): *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press., pp. 451-470.
- Manuel, S.Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 88: 1286-1298.

- Oh, E. 2008. Coarticulation in non-native speakers of English and French: An acoustic study. *Journal of Phonetics*, 36: 361-384.
- Öhman, S. E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39: 151-168.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J. and Hopper, P. (eds): *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, pp. 137-157.
- Recasens, D. 1985. Coarticulatory patterns and degree of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, 28 (2): 97-114.
- Recasens, D., Fontdevila, J. and Pallarès, M. D. 1995. Velarization degree and coarticulatory resistance for /l/ in Catalan and German. *Journal of Phonetics*, 23: 37-52.

# Rhythm and stress intervals in Greek and Russian

Antonis Botinis<sup>1</sup>, Marios Fourakis<sup>2</sup> and Olga Nikolaenkova<sup>1</sup>

<sup>1</sup>Department of Linguistics, University of Athens, Greece

<sup>2</sup>Department of Communicative Disorders, University of Wisconsin-Madison, USA

## Abstract

This paper presents the results of a pilot study of stress interval durations in Greek and Russian. Native speakers of each language were recorded producing utterances in which the number of syllables between stressed syllables varied. The results showed a noticeable tendency towards isochrony in Greek but not in Russian, according to which there is an inverse relation between the number of syllables and syllabic durations in produced utterances. Second, lexical stress has a lengthening effect in Greek but hardly in Russian, while focus has no effect in either language.

Key words: rhythm, focus, stress group, isochrony, duration, Greek, Russian

## Introduction

The term “stress group” is used in descriptions of the temporal properties of a language and is defined as a speech unit consisting of a stressed syllable and any unstressed syllables that may follow up to, but not including, the next stressed syllable (Pike 1946, Abercrombie 1967, Grønnum 1998). Languages are classified as either stress-timed in which stress groups and thus stress intervals tend to be isochronous such as English and Russian, or as syllable-timed in which syllables recur at regular intervals, such as French and Spanish. Absolute isochrony has not been reported for any language and the main question is about relative isochrony, which may vary between different languages with different prosodic structures.

Despite the stressed-timed and syllable-timed taxonomy of languages, acoustic evidence is basically ambiguous, leading some phoneticians to consider isochrony as a perceptual rather than a production related phenomenon (e.g. Lehiste 1977, Dauer 1983). This pilot experiment examined syllable and stress-group durations, as a function of different focus conditions in Greek and Russian.

## Experimental methodology

The speech material consists of four test sentences with a varying number of syllables per stress group for each test sentence, i.e. 1 to 4, produced twice by two Athenian and two Saint Petersburg female speakers, in their twenties, with focus in different places and at normal tempo (Table 1.). The speech material was directly recorded into a computer disc and measurements were carried out with the Praat software package.

Table 1. Test sentence syllabic sequences with different stress intervals in Greek and Russian (broad phonetic transcription and free translation).

Greek	Russian
i me.la.'ni 'ma.lo.ne ti 'ma.na mu Melany was scolding my mother.	ma.'ri 'my.la u.nix ix.man.da.'ri.ny Mary was washing at theirs their mandarins.
i me.'li.na 'ma.lo.ne ti 'ma.na mu Melina was scolding my mother.	ma.'ri.na 'my.la u.nix ix.ma.'li.nu Marina was washing at theirs their raspberries.
i me.'li.na mu 'ma.lo.ne ti 'ma.na Melina was scolding mother.	ma.'ri.na lo.'ma.la u.nix ma.'li.nu Marina has broken at theirs raspberries.
i me.'li.na mu ma.'lo.ni ti 'ma.na Melina is scolding mother.	ma.'ri.na na.lo.'ma.la ix.ma.'li.nu Marina has broken their raspberries.

## Results

Figure 1 shows the durations of stressed and unstressed syllables in Greek and Russian. The difference between stressed and unstressed syllable durations was significant for Greek ( $t(115)=5.758$ ,  $p<.01$ ) but not so for Russian ( $t(118)=1.474$ ,  $p>.05$ ). It should be noted (cf. the transcribed Russian sentences above) that the unstressed syllables used in Russian were more complex than the ones used for Greek. Even so, their durations do not exceed those of the stressed syllables.

Figure 2 shows the average durations of syllables as a function of number of syllables in the stress group in Greek and Russian. The differences in average syllable durations between stress groups were highly significant in Greek ( $F(3,113)=7.157$ ,  $p<.01$ ) but not so for Russian ( $F(3,116)=1.830$ ,  $p>.05$ ).

Figure 3 shows the durations of stress groups as a function of the number of syllables in Greek and Russian. Stress group durations increase as syllables are added to the group and this effect is significant for both languages. Analysis of variance showed a significant effect of number of syllables ( $F(3,87)=452$ ,  $p<.01$ ) and a significant effect of language ( $F(1,87)=44$ ,  $p<.01$ ). Stress groups were longer in Russian than in Greek. There was no interaction between number of syllables and language.

Figure 4 shows average syllable durations as function of focus in Greek and Russian. Syllables were longer in Russian than in Greek ( $F(1,231)=38$ ,  $p<.01$ ) but focus did not affect their durations ( $F(2,231)=.139$ ,  $p>.05$ ) and there was no interaction with language.

Figure 5 shows stress group durations as a function of focus for both languages. There were no significant effects or interactions.

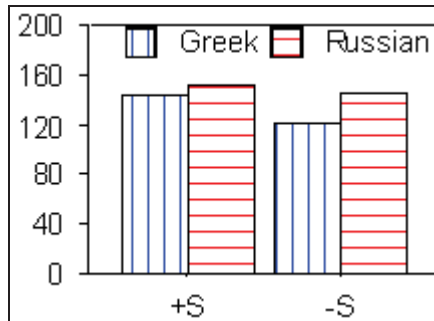


Figure 1. Average syllable durations as a function of stress (+stress/-stress) in Greek and Russian.

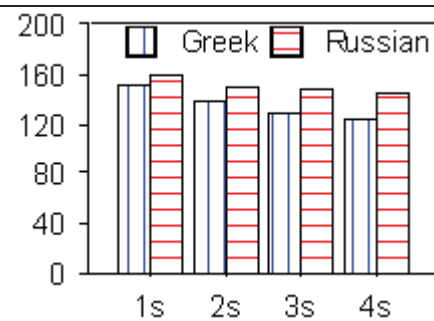


Figure 2. Average syllable durations as a function of syllables per stress group in Greek and Russian.

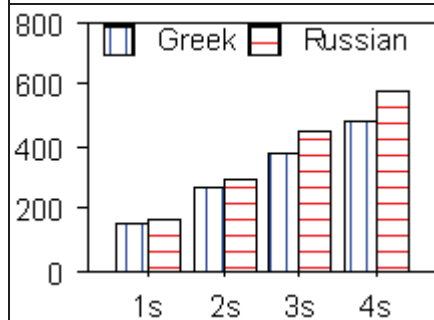


Figure 3. Average stress group durations as a function of syllables per stress group in Greek and Russian.

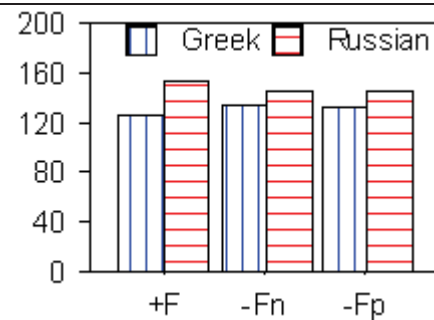


Figure 4. Average syllable durations as a function of focus (+focus/-focus neutral/-focus post position) in Greek and Russian.

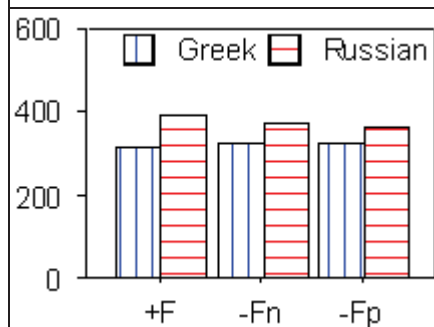


Figure 5. Average stress group durations as a function of focus (+focus/-focus neutral/-focus post position) in Greek and Russian.

### Discussion and Conclusions

A main finding of the present investigation is that the different focus conditions did not significantly affect the average durations of syllables and stress groups. This finding agrees with earlier research which found no significant durational effects of focus on either segments or syllables (e.g. Botinis 1989, Fourakis et al. 1999, Botinis et al. 2002). Russian seems to behave like Greek in this respect, according to the present results, which should however be further corroborated before conclusions are drawn.

Russian has been traditionally classified as a “stress-timed” language whereas Greek is reported as an “unclassified” language (see e.g. Dauer 1983). However, Greek and Russian do have similar rhythmic structures (see especially Fig.3 ), at least with reference to stress intervals presented in this paper, which is in accordance with Dauer’s conclusions that there is no clear-cut distinction between stress-timed and syllable-timed languages as most analysed languages show some degree towards isochrony. Still, both Greek and Russian show minimal isochrony for any of these languages to be taxonomised as a stress-timed language.

### References

- Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Botinis, A. 1989. *Stress and prosodic structure in Greek*. Lund: Lund University Press.
- Botinis, A., Bannert, R., Fourakis, M. and Pagoni-Tetlow, S 2002. Crosslinguistic segmental durations and prosodic typology. *Proceedings of Speech Prosody 2002*, 183-186. Aix-en-Provence, France.
- Dauer, R. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.
- Fourakis, M., Botinis, A. and Katsaiti, M. 1999. Acoustic characteristics of Greek vowels. *Phonetica* 56, 28-43.
- Grønnum (Thorsen), N. Intonation in Danish. In Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems*, 131-151. Cambridge: Cambridge University Press.
- Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5, 253-263.
- Pike, K.L. 1946. *The intonation of American English*. Ann Arbor: University of Michigan Press.



# Investigations of speech segmentation: addressing the writing bias in language research

Victor J. Boucher and Annie C. Gilbert

Laboratoire de sciences phonétiques, Université de Montréal, Canada

## Abstract

We address the problem of defining universal processes of speech segmentation in view of criticisms that conceptual linguistic units derive from western writing. A synthesis of our recent experimental studies is presented bearing on processes of serial-order and rhythmic grouping. First, on how serial-order operates, we use EMG and speech-motion data to show that “consonant-vowel” orders reflect contraction-relaxation cycles (not separate segments). Second, on how sequences of sounds come to form units in language learning, we discuss behavioural data suggesting a link between rhythm groups in speech and grouping effects on memory of speech sounds. Preliminary EEG data is then presented to substantiate the view of an on-line parsing of rhythm groups with effects on memory traces of lexemes.

Key words: speech, segmentation, universals, physiology, history of linguistics

## The writing bias and the need to define segmentation

The importance of defining processes of segmentation can be weighed by considering the weak validity of formal linguistic units. Historically, though principles like “phonological distinctiveness” played an essential epistemological role in defining features, one finds no principle forcing the view that features occur in “bundles” as in letters of IPA, or that meaningful elements (morphemes) occur in “word” units as in text. In fact, several critics have warned that concepts of *phoneme* and *word* implicitly refer to European writing and orthographic codes (e.g. Coulmas, 1989; Linell, 2005). But while critics point to the fundamental problems arising from a writing bias in conceptualizing units, little research is devoted to defining the processes by which features are ordered and learned in multisyllabic groups. On these issues, we provide a synthesis of our recent work on segmentation, which bears centrally on the processes underlying serial order and rhythmic grouping.

## Defining the process of serial-ordering

Using IPA to represent speech can foster a conception that serial order follows letter-like “segments”. Such conception pervades not only linguistic theory but also neural network and speech-motor models. All assume that, in producing a syllable like /pa/ neural influx to articulators, as can be observed by EMG, is organized serially with closer muscles being activated before

openers in line with notions of consecutive consonant and vowel segments. However, Boucher (2008) showed that consecutive activation may not occur. Figure 1 from that study illustrates the activity of the main muscles involved in aperture motions of the lips and jaw. Note that, in creating close-open cycles of the lips while clenching the teeth, activity appears for lip-openers (depressor labii m.). Conversely, in creating close-open cycles of the jaw with lips sealed, activity appears for jaw openers (the anterior digastric m.). However, in creating close-open cycles in speaking /papapa/, activity of openers is not present. This questions the notion of separate activation by reference to [+cons] and [-cons] phonemes and suggests instead a serial ordering where passive factors such as tissue elasticity affects openings.

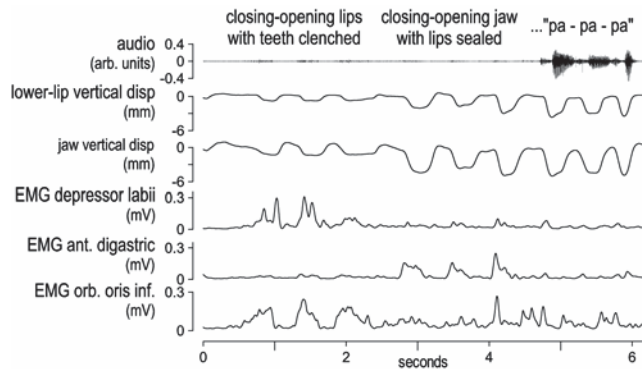


Figure 1. EMG for close-open cycles of the jaw and lips during non-speech and speech. Adapted from Boucher, 2008.

This latter suggestion was investigated by hypothesizing that, if opening motions reflect passive spring-like behavior, then Hooke's law would apply: force applied in compressing a spring leads to an opposite elastic force, which can move a mass at a speed and distance proportional to system constants. We took oral pressure and bilabial compression during closing as indices of applied force and predicted that velocity and distance of lip and jaw opening would be linearly related to applied force. Strain gages were used to monitor lip and jaw motion and 3 subjects were asked to produce series *papa...* and *baba...* with increasing loudness so as to emphasize changes in lip compression and oral pressure. Figure 2 presents an example of the results: both force indices showed such linear relations with speed and range of opening for *ba* and *pa* series, as predicted. In short, close-open cycles of motion in articulators may not reflect serial activation for assumed [+cons] and [-cons] phonemes as currently assumed in motor theories. Instead serial control follows cycles of contraction and relaxation with opening motion being largely driven by passive factors. One should also note that this view of serial-ordering conforms to the perception of speakers who do not know alphabet writing. These speakers can easily count "syllable cycles" but not letter-like phonemes (see Boucher, 1994 in Boucher, 2008).

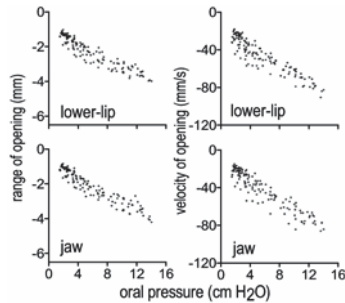


Figure 2. Example of linear relationships between an index of closing force (i-o pressure) and both the speed and range of opening motion, suggesting elasticity effects. Adapted from Boucher, 2008.

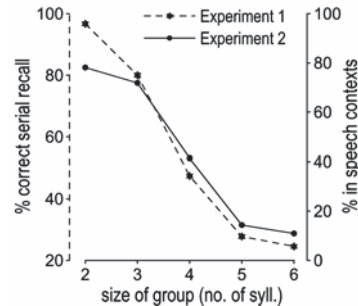


Figure 3. Effects of size of groups on recall (*Experiment 1*) and frequency of occurrence of group sizes in speech contexts (*Experiment 2*). Adapted from Boucher, 2006.

### Defining the grouping process in learning language forms

The above defines a basic process of serial ordering. However, specific sequences of speech sounds come to consolidate as forms in language learning. This implies a capacity to parse and hold in memory a given chunk of speech. What is the extent of such parsing and are there any marks of this process in speech? One clue appears when speakers recall novel series of syllables or digits. In such tasks, groupings arise, and it is known that recall is enhanced when lists of items are presented in groups of 3 or 4. What is intriguing is that there is also a tendency in speech to create rhythms that do not exceed 3 or 4 cycles (Boucher, 2006). Considering this coinciding “size-effect,” we explored the link between grouping effects on recall and rhythm groups (RGs) in speech by two experiments involving 40 French speakers.

In *Experiment 1* subjects had to recall heard series of 7 non-sense syllables. Two sets of stimuli were constructed from monotone speech. The first presented set were arrhythmic series of unstressed syllables (*s*); the second set were series with long syllables (*S*), one placed at the end of a sequence and one placed at varying internal positions so as to create rhythms of 2 to 6 syllables (i.e. *s S s s s s S*, *s s S s s s S*, *s s s S s s S*, etc.). In *Experiment 2*, sentences with subject and verb-complement phrases of 2 to 6 syllables were visually presented. Subjects had to say a context and then repeat it twice from memory, once normally and once using glottal stops to mark rhythm. For instance, we used sentences with short and long compound names (*Pierre-Paul part mercredi matin*; *Marie-Antoinette part mercredi matin*) to determine how rhythmic grouping operates when constituents vary in length. The results of *Experiment 1* showed that only groups not exceeding 4 syllables significantly enhance memory compared to arrhythmic

series. More importantly, as seen in Figure 3, the decreasing benefit of long groups in recall appeared to correlate narrowly with the decreasing frequency of long RGs observed in the speech contexts of *Experiment 2*.

While the above results support the hypothesis of a link between groupings that benefit memory and RGs in speech, they do not directly demonstrate an on-line parsing of the groups in listening to utterances. On this problem, a preliminary study using the EEG technique of evoked potentials was conducted (Gilbert et al, 2008). Controlled contexts were presented to a listener consisting of 50 utterances bearing RGs of 4 and 5 syllables *within* (and not at the boundaries of) intonation groups, and these were presented along with distractor utterances. The task consisted of responding to the presence/absence of a given lexeme in the utterance. The results revealed known "closure positive shifts" evoked by length marks of RGs while the subject was listening to the utterances. Also, decreases in N400 were found for target lexemes previously presented in the short RGs, suggesting that, compared to 5-syllable RGs, shorter groups benefit memory traces of lexemes. These results confirm an on-line parsing of heard utterances by reference to RGs and suggest that group-size affects memory of forms within the groups.

### **Prospective conclusion**

It should be noted that the above processes of segmentation are language-independent. They can give rise to forms which may be variably termed and have varying roles in different language systems (e.g. syllables, mora, lexemes, concatenations, etc.). However, applying language-relative terminology and analyzing transcriptions that assume letter-like segments and word divisions may not capture the universal process at the source of the forms.

### **References**

- Boucher, V. J. 2006. On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language and Speech*, 49, 495-520.
- Boucher, V. J. 2008. Intrinsic factors of cyclical motion in speech articulators. *Journal of Phonetics*, 36, 295-307.
- Gilbert, A. C., Boucher, V. J. and Jemel, B. 2008. Size of rhythm-groups affects the memory trace of heard words in utterances. In P. A. Barbosa, S. Madureira and C. Reis (Eds.), *Proc. of the Speech Prosody 2008*, 379-382. Campinas, Brazil.

## Two sources of voicing neutralization in Lithuanian

Rebeka Campos-Astorkiza

Department of Spanish and Portuguese, The Ohio State University, USA

### Abstract

This study compares two processes that result in voicing neutralization in Lithuanian: regressive voicing assimilation in obstruent clusters and final devoicing of obstruents. Acoustic data is analyzed to assess the behaviour of three acoustic cues to obstruent voicing (i.e. closure and preceding vowel duration and voicing during closure) in both neutralizing environments. The results show that, although both processes result in incomplete voicing neutralization, they use the acoustic cues differently. This suggests that final devoicing and voicing assimilation have different acoustic realizations, supporting their analysis as two different processes.

Key words: neutralization, voicing assimilation, final devoicing

### Introduction

Previous experimental studies on voicing neutralization have primarily focused on final devoicing (e.g. Dinnsen and Charles-Luce 1984, Port and Crawford 1989). Their main finding is that neutralization may be incomplete and that several factors, including semantic and pragmatic, play a role in determining the degree of neutralization. However, less attention has been paid to another potential source of voicing neutralization, namely, voicing assimilation (but see Slis 1986, Charles-Luce 1987, 1993, Burton and Robblee 1997). This study compares neutralization through final devoicing and voicing assimilation in order to evaluate whether both processes lead to a similar degree of neutralization. Lithuanian is used for this purpose because it displays both processes. Word final obstruents undergo devoicing. Regressive voicing assimilation occurs in obstruents clusters, such that the voicing of the last member determines the voice realization of any preceding obstruents. Sonorants do not participate in the process. They do not trigger or undergo voicing assimilation (Mathiassen 1996). Thus, an acoustic experiment was designed to analyze voicing neutralization in Lithuanian as manifested in three temporal intervals that have been previously established as cues to obstruent voicing: preceding vowel duration, closure duration and amount of voicing during closure. More precisely, longer vowel duration and voicing during closure are associated with voiced obstruents, and longer closure duration is correlated with voiceless ones.

### Methodology

Two sets of nonsense words were constructed according to Lithuanian phonotactics: one for voicing assimilation and another for final devoicing. Stimuli in the voicing assimilation condition consisted of bisyllabic words of the form tV1C1C2a, with stress on the first syllable. V1 could be any of the long vowels from the Lithuanian inventory /i:, e:, æ:, a:, o:, u:/. The medial cluster C1C2 could be /k3/ and /g3/ for the assimilatory contexts, and /kʃ/ and /g3/ for the non-assimilatory contexts (i.e. where underlying stop voicing is expected to match its surface realization). Some sample tokens are /ti:k3a/, /ti:g3a/, /ti:kʃa/ and /ti:g3a/. The stimuli for the final devoicing condition were also bisyllabic nonsense words of the shape datV1C1, with stress on the last syllable. V1 could be any of the long vowels in Lithuanian and the last consonant C1 was /k/ or /g/. Some sample tokens are /dati:g/ and /dati:k/. Standard Lithuanian orthographical representations were employed to mark voicing differences. The relevant words were inserted in the carrier sentence *Sakyti \_\_ negalima* “To say \_\_ is not allowed”. Each stimulus was repeated 8 times. Each block of 36 sentences was pseudo-randomised. Five native speakers of standard Lithuanian, one male and four females, were recorded. The sentences were displayed on a computer screen and speakers were asked to read each sentence in a colloquial style. They were cued for each sentence to keep the rhythm constant. Before the actual experiment, speakers were given some practice tokens.

The data was analyzed using synchronized waveforms and spectrograms to measure the duration of the vowel, stop closure and voicing during closure. The vowel was measured from the onset of the first glottal pulse to the offset of the last one in the waveform, before voiceless stops. Preceding voiced stops, the end of the vowel was determined by a drop in amplitude and a change in waveform shape, or where the formant structure ended. Stop closure duration was measured from the end of the preceding vowel to the closure release or the beginning of frication, when there was no clear release. Beginning of frication corresponded with the start of aperiodic energy. The duration of voicing during closure was calculated from the beginning of the closure to the end of the last visible glottal pulse.

### Results

For the assimilation condition, three-factor (vowel quality, underlying stop, following fricative) repeated measures ANOVAs were performed for each temporal interval. For the final devoicing condition, two-factor (vowel quality, underlying stop) repeated measures ANOVAs were carried out for each temporal interval. Repetitions per test word were averaged within subjects, and the significance level was set at  $p < .05$ . Only main effects of

underlying voicing and following fricative (in the assimilation condition) are discussed. Also, none of the analyses showed an interaction between vowel quality and underlying stop, indicating that all vowel qualities behave similarly with respect to underlying stop voicing for all dependent variables and conditions.

Let's begin with the assimilation condition. For vowel duration, there is a significant effect of underlying stop ( $p=.035$ ) so that vowels before underlying voiceless stops are shorter than before voiced ones, and of following fricative ( $p<.001$ ), so that vowels are longer in the voiced assimilation context than in the voiceless. There is a significant effect of underlying stop ( $p=.02$ ) and of following fricative ( $p=.01$ ) on closure duration. The closure for underlying voiceless stops is longer than for voiced ones, and closure duration is longer in the voiceless assimilation context than in the voiced. As for voicing duration, there is a main effect of underlying stop voicing ( $p=.035$ ) and of following fricative ( $p<.001$ ). Voicing is longer for underlying voiced stops than for voiceless ones. Voicing lasts longer in the voiced assimilation context than in the voiceless. There is no significant interaction between underlying stop and following fricative for any of the intervals.

Moving on to the final devoicing condition, the results indicate that underlying stop has a significant effect on vowel duration ( $p=.02$ ) and closure duration ( $p=.036$ ) but not on the amount of voicing during closure.

Table 1. Mean vowel, closure and voicing duration (ms) across speakers and vowel qualities for the assimilation condition.

	Following /ʒ/			Following /ʃ/		
	Vowel	Closure	Voicing	Vowel	Closure	Voicing
Underlying /g/	167	52	49	150	58	15
Underlying /k/	163	56	38	144	60	0

Table 2. Mean vowel, closure and voicing duration (ms) across speakers and vowel qualities for word final obstruents.

	Vowel duration	Closure duration	Voicing duration
Final /k/	145	91	0
Final /g/	166	77	33 (st.dev.=35 due to one subject)

Finally, in order to obtain a direct comparison between final devoicing and voicing assimilation, a separate two-factor (vowel quality, underlying stop) repeated measures ANOVA on each temporal interval was conducted for those tokens where voicing assimilation resulted in devoicing (/gʃ/ & /kʃ/). The results show that underlying stop is a significant factor only for



voicing during closure ( $p=.044$ ). For vowel and closure duration, it fails to reach statistical significance indicating that these two temporal intervals are similar for the stops in the /gʃ/ & /kʃ/ tokens.

### **Discussion and conclusion**

The results suggest that voicing neutralization from either assimilation or devoicing may be incomplete in Lithuanian. This is line with previous findings for other languages. More interestingly, our data suggest that neutralization due to final devoicing and due to voicing assimilation have different acoustic realizations, as shown by the way in which each acoustic correlate is employed in each case. Charles-Luce (1987) also found that neutralization applied to a different degree depending on whether it was the result of assimilation or final devoicing in Catalan. In the present study, at least for voiceless assimilation, neutralization seems to apply to a greater extent than in final devoicing. Only one acoustic cue differentiates underlying voiceless and voiced stops in this context, as opposed to final devoicing where two acoustic cues differentiate the voicing contrast in the surface realization.

These experimental results lend some support to the analysis of voicing assimilation and final devoicing as two distinct processes that lead to two different representations of voicing and thus, two different realizations. More precisely, I argue for an account of (incomplete) voicing neutralization in Lithuanian which distinguishes between final devoicing and voicing assimilation, as opposed to a unifying analysis, whether phonological or phonetic, for both processes, for instance, through a single feature-value change or unspecification of voicing features.

### **References**

- Burton, M.W. and Robblee, K.E. 1997. A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25, 97-114.
- Charles-Luce, J. 1987. An acoustic investigation of neutralization in Catalan. Ph.D. dissertation, Indiana University.
- Charles-Luce, J. 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50, 28-43.
- Dinnsen, D.A. and Charles-Luce, J. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12, 49-60.
- Mathiassen, T. 1996. A short grammar of Lithuanian. Slavica Publishers.
- Port, R. and Crawford, P. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17, 257-282.
- Slis, I. 1986. Assimilation of voice in Dutch as a function of stress, word boundaries, and sex of speaker and listener. *Journal of Phonetics* 14, 311-326.

# **Stress assignment in Brazilian Portuguese: a usage-based approach**

Maria Cantoni

Faculty of Letters, Federal University of Minas Gerais, Brazil

## **Abstract**

The major debate on primary stress assignment in Brazilian Portuguese (BP) concerns whether it is lexically given or predictable by a set of principles. This paper presents a contribution to this debate by suggesting a usage-based approach to stress assignment in BP. It is argued that stress assignment is better accounted for as the result of generalizations over exemplars, and these generalizations are responsible for the main tendencies regarding stress assignment in the language.

Key words: stress. Brazilian Portuguese. usage-based phonology. exemplar model

## **Introduction**

The major debate on primary stress assignment in Brazilian Portuguese (henceforth BP) concerns whether it is lexically given or predictable by a set of principles. This paper presents a contribution to this debate by suggesting a usage-based approach to stress assignment in BP. In the first section, an outline of usage-based models is presented. The second section addresses the main claims about BP stress, which are analysed according to a usage-based approach in the third section. The fourth section presents final conclusions.

## **Usage-based models**

Usage-based models claim that linguistic systems are made out of instances of use, and they contrast with formalistic frameworks for assuming that in the storage of information, a large amount of data is involved, and with great redundancy (Langacker 1987). This paper relies mostly on three relatively recent usage-based approaches that focus on the sonority dimension: Usage-Based Phonology (Bybee 2001), the exemplar model proposed in Pierrehumbert (2001), and Probabilistic Phonology (Pierrehumbert 2003), as well as Langacker's proposal. The models adopted in this paper are briefly outlined below. They share the assumption that there is no clear-cut separation between lexicon and grammar. Morphological and syntactical patterns are viewed as schemas that emerge from real instances of linguistic use. Therefore, frequency of type and token plays an important role, having an impact on productivity of patterns and on lexical strength, respectively. Linguistic units of storage and processing, i.e., the exemplars, are connected in relational networks based on semantic and phonetic similarity. Such

networks, from which general schemas emerge, operate on the basis of probabilistic parameters, yielding a stochastic distribution.

### Brazilian Portuguese primary stress

Regarding BP stress assignment, most traditional analyses (cf. Bisol 1992; Cagliari 1999; Câmara Jr. 1970; Lee 1995; Massini-Cagliari 1992; Wetzels 2006, *inter alia*) agree upon two major issues: (1) stress falls on a final three-syllable window; (2) penultimate stress is the most recurrent one. However, there is divergence over a number of related issues, the most important one concerning the differences between stress assignment in nominal and verbal morphology. Whereas the former usually presents penultimate stressed forms, mostly with open syllables, the latter typically presents penultimate stress, mostly with a closed syllable. Further, verbal morphology presents postonic nasal vowels, which do not usually occur in nominal morphology.

It can be argued that stress is lexically contrastive in BP on the basis of minimal pairs such as *sábia* “wise (fem.)”, *sabía* “used to know (1<sup>st</sup>/3<sup>rd</sup> sing.)” and *sabiá* “song-thrush” (stressed syllables in bold). However, it is not simple to find minimal pairs in the same word class—e.g., *cara* “face” vs. *cará* “yam”; *comeram* “ate (3<sup>rd</sup> pl.)” vs. *comerão* “will eat (3<sup>rd</sup> pl.)”.

In BP, stress is acoustically correlated to an increase in the relative duration and intensity (Massini-Cagliari 1992), and given a word the stressed syllable is to some extent fixed. Actually, in BP phenomena involving stress shift (e.g. *fluido* ~ *fluidido*) are rather rare. Another phenomenon related to stress, in which the stress pattern of a word is changed due to loss of sound material (e.g. *abóbora* ~ *abobra*), seems to be more productive.

### A usage-based analysis of stress in BP

The present study relies on the idea that segmental and prosodic patterns interact with morphology in a network model in order to configure the Prosodic Grammar (Bybee 2001). Traditional approaches to BP stress assignment treat generalizations as the ultimate cause of regularity, instead of the result of routinized patterns and categorization processes, as I rather assume. A similar analysis is proposed by Farrell (1990), who applies Langacker’s cognitive model to account for Spanish stress assignment in non-verbs. In my proposal, as in his, lexical stress integrates higher-level schemas, such as morphology and phonotactics, being part of the information conveyed by tokens of experience and stored in representation.

It is argued that if stress is lexically given then a more comprehensive analysis for productiveness of stress patterns is met, both in verbal and non-verbal morphology. To support this claim, a general statistical analysis of stress patterns in verbs vs. non-verbs is presented. Such analysis is based on

information from ASPA ([www.projetoaspa.org](http://www.projetoaspa.org)), which is a statistical database for BP sound patterns. Table 1 presents type frequency in each of the three stress positions, for verbs and non-verbs. Penultimate stress has a significantly higher type frequency, and this can explain why this pattern is more productive than the other two, as is frequently claimed in the studies on BP stress previously mentioned. This is corroborated by the general trend in the stress location of neologisms (cf. those presented in Alves 1994), mostly penultimate-stressed. Token frequency differences among the same groups are not statistically significant and are not addressed in this analysis.

Table 1. Type frequency of stress patterns ( $\chi^2$ : 5642.69,  $p < 0.0001$   $df=2$ ).

Stress Position	Antepenultimate		Penultimate		Penultimate	
Verbs	482	1.2%	27,730	71.4%	10,617	27.3%
Non-verbs	11,389	14.6%	52,271	67.1%	14,200	18.2%

Figure 1 represents prosodic schemas emerging from network relations between exemplars, for verbal forms. The particularities identified above between stress patterns of verbs and non-verbs can be attributed to different mechanisms of storage and management of the lexicon (Bybee 1985).

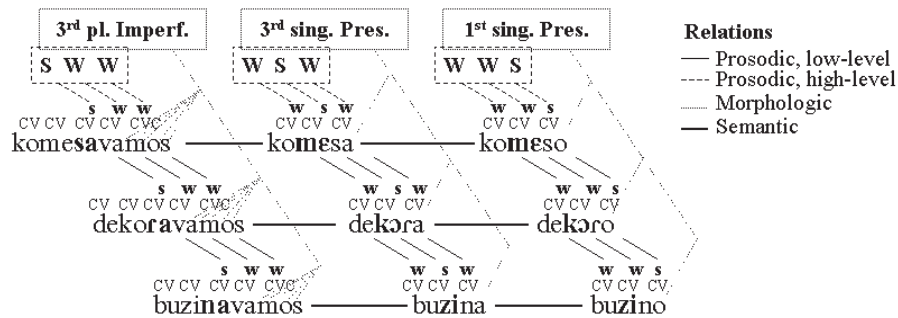


Figure 1. Emergent Prosodic Structure and high-level schemas in verbs.

In BP, verbs present a complex and close-knit morphologic network arising from inflection, that encodes voice, aspect, tense, mood, person and number agreement. Comparatively, inflectional relations in non-verbs are reduced and weaker, encoding only number and sometimes gender. It can be argued that complex and dense inflectional networks are responsible for a greater degree of integration in higher-level schemas. If this is so, it can be easily demonstrated that stress schemas arisen from verbs (but not from non-verbs) tend to be attached to the morphological schemas that parallel them.

### Concluding remarks

A number of issues must be raised from the analysis presented in this paper, such as: a) how a statistical analysis of speech offers generalizations related to general tendencies in the language involving sounds and morphological information; b) the relationship between stress assignment and syllabic patterns. It is claimed in this paper that primary stress in BP is lexically specified, and this supports the idea that redundant information is present in mental representations. It is thus argued that stress assignment is better accounted for as the result of generalizations over exemplars, and these generalizations are responsible for the main tendencies related to stress assignment in BP. Further research intends to develop some of the ideas presented in this paper aiming at a probabilistic modeling of stress.

### Acknowledgements

I am indebted to Thaïs Cristófar, for comments and suggestions, and to Leonardo Almeida, for the information from ASPA database. I would also like to thank Joan Bybee and Patrick Farrell for sending me relevant bibliographical references. The present work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq/Brasil.

### References

- Alves, I. 1994. Neologismo: criação lexical. São Paulo, Ática.
- Bisol, L. 1992. O acento e o pé métrico binário. *Cadernos de Est. Ling.* 22, 69-80.
- Bybee, J. 1985. *Morphology: a study of the relation between meaning and form.* Philadelphia, John Benjamins.
- Bybee, J. 2001. *Phonology and Language Use.* Cambridge, Cambridge Univ. Press.
- Cagliari, L. 1999. *Acento em Português: estudos sobre as regras de atribuição de acento em português.* Campinas, Author's edition.
- Câmara Jr., J. 1970. *Estrutura da língua portuguesa.* Petrópolis, Vozes.
- Farrell, P. 1990. Spanish stress: A cognitive analysis. *Hispanic Linguistics* 4, 21-56.
- Langacker, R. 1987. *Foundations of cognitive grammar.* Stanford, Stanford Univ.
- Lee, S. 1995. *Morfologia e Fonologia Lexical no Português do Brasil.* Ph.D. diss., IEL-UNICAMP.
- Massini-Cagliari, G. 1992. *Acento e ritmo.* São Paulo, Contexto.
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. and Hopper, P. (eds.) 2001, *Frequency and the emergence of linguistic structure*, 137-157. Amsterdam, John Benjamins.
- Pierrehumbert, J. 2003. Probabilistic Phonology: discrimination and robustness. In Bod, R., Hay, J. and Jannedy, S. (eds.) 2003, *Critical Introduction to Phonology*, 177-228. Cambridge MA, MIT.
- Wetzels, W. 2006. Primary Word Stress in Brazilian Portuguese and the Weight Parameter. *Journal of Portuguese Linguistics*, 5(2), 9-58.

# The temporal structure of professional speaking styles in Brazilian Portuguese

Luciana Castro<sup>1</sup> and João Antônio de Moraes<sup>1,2</sup>

<sup>1</sup>Laboratory of Acoustic Phonetics, Universidade Federal do Rio de Janeiro, Brazil

<sup>2</sup>Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

## Abstract

This study describes the temporal organization of speech collected from politicians, religious leaders and television news anchors, individuals for whom speech represents an essential component of their professional identity. Analysis of duration measurements of the recorded speech suggests that the temporal differences observed between the speaking styles are statistically significant.

Key words: prosody, speaking style, temporal organization of speech

## Introduction

The relation between prosody and the characterization of a professional speaking style is noted by Léon (1993), who observes that different situations (public speaking, reading, theater, etc.) result in different speaking styles which over time become permanently associated with particular professions. This study investigates whether the temporal structure of speech differs significantly among the professional speaking styles of the politician, the religious leader, and the news broadcaster for Brazilian Portuguese.

## Method

The analysis is based on the professional speech of six adult speakers: two presenters of TV news, two TV ministers (religious leaders), and two politicians. The speakers, ranging in age from 35 to 55, are native speakers of Brazilian Portuguese with no apparent speaking disorders.

Each recording consisted of two minutes of speech captured during the exercise of each speaker's profession: the TV news presenters in the studio, the politicians in the senate, and the religious leaders in their studio/church. All of the recordings were captured directly from Brazilian broadcast television. From observing the video, it was apparent that none of the speakers was reading directly from a text, though it is assumed that the TV news anchors were working with the support of a teleprompting machine.

The recordings, made on a laptop computer connected to the audio output of an analog television, were digitized using a sampling rate of 22 kHz and analyzed using the acoustic analysis software Speech Analyzer.

The recordings were transcribed orthographically, and then silent and filled pauses in the speech were identified through a combination of perceptive analysis (listening) and the visual inspection of the spectrograms. As a rule, any interruption of more than 40 milliseconds in the flow of speech was considered a pause. The result of this process was the segmentation of the speech in two units: pauses and speech sequences.

Based on this segmentation, it was possible to derive the following metrics: the total time of the pauses (TP), the number of pauses (NP), the average pause duration (AP), the pause time per syllable (TP/syl), the speech rate (SR), and the articulation rate (AR).

## Results and discussion

### Types of pauses: silent vs. filled

Filled pauses are completely absent in the TV news speaking style, whereas in the sermon and in the political speech, 3.9% and 10.6% of the pauses, respectively, are filled pauses, suggesting a greater degree of spontaneity in the speech of the politician as compared to the other professional speaking styles.

### Number and duration of pauses

The TV news speaking style exhibited the fewest number of pauses, followed by the political and religious styles, as can be seen in Figure 1.

The average pause duration, in principle independent of the number of pauses, follows exactly the same order, with the TV news speaking style having the shortest average pause duration, followed by the political speaking style, and finally the religious speaking style (Figure 1). As such, the total pause time (TP, not shown here), being the product of the number of pauses and the average pause time, exhibits the same tendencies (to a greater degree) present in each of the component variables.

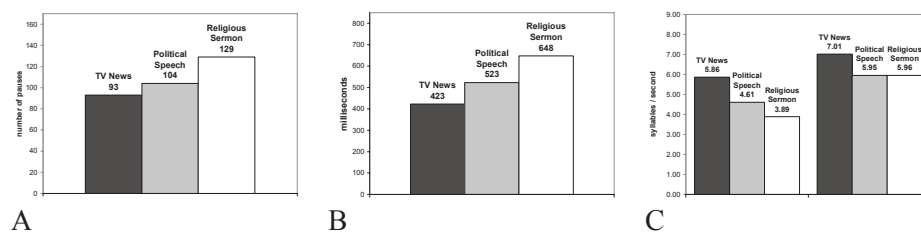


Figure 1: Total number of pauses (A), average pause duration (B), and speech and articulation rate (C) for each speaking style.

Previous studies (Sabin et al. 1979) have observed that read speech employs fewer pauses than spontaneous speech, which may explain the



lower number of pauses observed here in the speech of the TV news broadcasters. Similarly, Goldman-Eisler (1968) reports that the total pause duration [TP] for “descriptive” speech is half that of “interpretive” speech. One factor that may contribute to this observed difference in the speech of religious leaders and politicians is their use of pauses for dramatic effect (Duez, 1991).

### Speaking rate

Figure 1 (above) also presents the results of measuring both the overall speech rate and the articulation rate, which considers only the segments of the recording labeled as “speech sequences” (that is, excluding the pauses).

The religious and political speaking styles were delivered at the same articulation rate (5.96 syllables/second), while the TV news was presented 17.6% faster (7.01 syl/sec). When the pauses are taken into account, overall speech rate inversely reflects the tendencies previously reported for total pause time, with the religious speaking style presenting the slowest speech rate at 3.89 syl/sec, followed by the political speaking style at 4.61 syl/sec and, finally, the TV news speaking style with the fastest speech rate at 5.86 syl/sec. Shevchenko and Uglova (2006) observed speaking rates of 5.1 syl/sec for TV news. Delgado-Martins and Freitas (1991) suggested that the speech of TV news broadcasters may be faster due in part to the severe time restrictions imposed on the delivery of the news, a factor that may also explain the reduced average duration of the pauses.

### Statistical analysis

The histograms in Figure 2 suggest that the pause durations of each speaking style do not obey a normal distribution. In order to evaluate the significance of the differences observed between the speaking styles, the non-parametric Kruskal-Wallis test, which does not assume a normal distribution, was applied. The test was applied using the statistical software package “R” and resulted in a p-value of  $2.294 \times 10^{-6}$  – that is, significant for  $\alpha = 0.01$ .

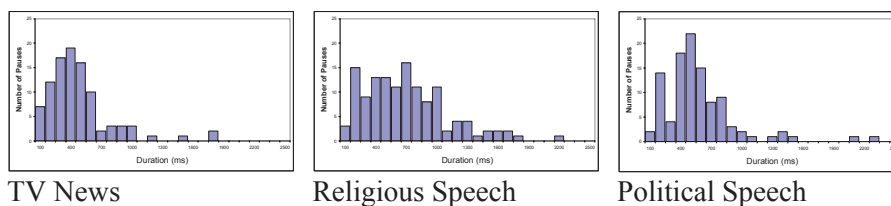


Figure 2: Histograms of pause durations for each speaking style.

Note that the religious speaking style, in addition to being the style with the highest pause duration, is also the style with the largest variance.



### Conclusions

The temporal characteristics observed in this study corroborate the findings reported in previous work in several regards. First, the semi-spontaneous speech of the religious leaders and politicians was characterized by a greater number and average duration of pauses and by a slower speaking rate than that of the TV news broadcasters. These results are in accordance with the idea that, in speech in which lexical and syntactic choices are the responsibility of the speaker in real time, the time spent in pauses suffers an increase (Sabin et al. 1979). Furthermore, a total absence of filled pauses was observed in the TV news speaking style, in contrast to the political and religious speaking styles.

A statistical analysis of the data revealed significant differences in the pause duration distributions of the three speaking styles studied here. However, due to the limited size of the corpus, it was not possible to infer a direct relation between the professional speaking style *per se* and the observed values. Specifically, it was not possible to reject the hypothesis that the differences observed are a result of the personal speaking style of each speaker, as opposed to the professional speaking style.

### References

- Delgado-Martins, M.R. and Freitas, M.J. 1991. Temporal structures of speech: reading news on TV. Proceedings of the ESCA Workshop, Barcelona, Spain.
- Duez, D. 1991. La pause dans la parole de l'homme politique. Paris, CNRS.
- Goldman-Eisler, F. 1968. Psycholinguistics: experiments in spontaneous speech. London and New York, Academic Press.
- Léon, P. 1993. Variation situationnelle et voix professionnelles. In: Précis de Phonostylistique. Paris, Nathan.
- Sabin, E; Clemmer, E; O'Connell, D; Kowal, S. 1979. A pausological approach to speech development. In: Siegman, A. and Feldstein, S. (eds.) Of speech and time. New Jersey, LEA Publishers.
- Shevchenko, T and Uglova, N. 2006. Timing in news and weather forecasts: implications for perception. Proceedings of Speech Prosody, Dresden, Germany.

# Prosodic perception of sentence types in Greek

Anthi Chaida  
Department of Linguistics, University of Athens, Greece

## Abstract

This study examines the relation between prosody and sentence-type perception in Greek, through a perceptual experiment using synthetic hum sound analogs of statements, polar/wh-questions and commands. The results indicate that sentence types may be partially identified through prosody. The identification rates were not very high, due to the nature of the synthetic stimuli. Wh-questions were easily perceived, while for the other sentence types falling boundary tones seem to be misleading.

Keywords: perception, sentence types, prosody, Greek

## Introduction

This experimental study investigates the effects of prosody on sentence type-perception in Greek, i.e. whether prosody can be a decisive factor for the perception of statements, polar/wh-questions and commands. Previous studies showed that naturally produced stimuli of all types were perceived with about 99% accuracy (Chaida 2005). This experiment is focused on artificial (hum) sound analogs, where all linguistic information, except prosody, is eliminated. In the absence of all lexical information, the distinction between sentence types is assumed to depend on the prosodic structure, especially in languages like Greek and Italian. Sentence types are associated with local and global tonal structures (e.g. Botinis 1998, 't Hart, 1998, Makarova 2001). Studies in Greek prosody have revealed several tonal characteristics for each sentence type (e.g. Baltazani 2007, Chaida 2007).




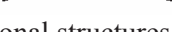
Sentence Type	Tonal structure	Boundary
STATEMENT		Low
POLAR QUESTION		Rise-Fall
WH-QUESTION		Rise
COMMAND		Low

Figure 1: Stylized tonal structures for sentence types in Greek.

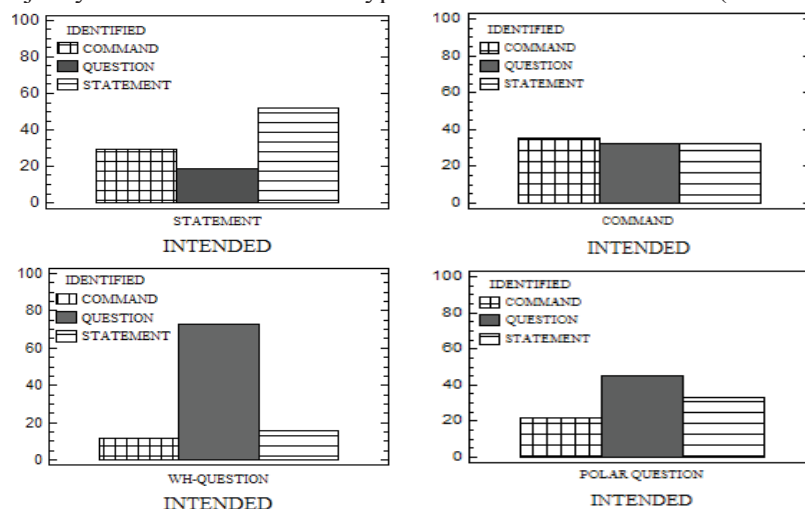
## Materials and Method

A set of utterances produced by 3 female speakers in their twenties was selected from a previously recorded corpus of Greek speech material (Chaida 2007). It included variations of the simple sentence ‘*O Manólis mazévi lemónja*’ (Manolis is picking lemons) and the complex ‘*O Manólis mazévi lemónja ce/ótan/parólo pu i María mirázi balónja*’ (Manolis is picking

lemons and/when/although Maria is distributing balloons), coordinated and subordinated, produced as statements, polar/wh-questions and commands, mainly by prosodical means and minor alterations (wh-word '*jati*' for wh-questions, and the imperative verb form for commands). This set of 20 recorded utterances was randomized 5 times. The perception test stimuli were processed with Praat 4.5.21, in order to create sounds with the algorithm described at Point Process: To Sound (hum) (modifying formants), so as prosody would be the only factor judged without any lexical influences. The modified stimuli were presented to 32 informants (male and female, 20-40 years old) through a program designed on C#. The experiment took place through earphones and the informants were instructed to categorize each stimulus (statement, question, command), and then provide an indication of how certain they were (1-6 scale: 1=least, 6=most certain).

## Results

The results are shown in Figures 2-5. In the first instance, it is evident that, although overall identification rates are not as high as might be expected, the majority of identified sentence types match the intended ones (35-72%).



Figures 2-5: Identification for intended statement, command, wh- and polar question productions.

Stimuli based on **statements** (Fig.2) were identified as intended 52.08% of the time, while 29.43% were identified as commands and 18.49% as questions. The mean certainty rating was 3.32 (out of 6), but the ANOVA table showed that certainty was not significant for statements ( $p > 0.05$ ,  $F = 1.33$ ). Stimuli based on **commands** (Fig. 3) were identified as intended 35.42% of the time, while 32.47% were identified as commands and 32.12% as statements. The mean certainty rating was 3.41, although certainty was

not significant for commands ( $p > 0.05$ ,  $F = 2.53$ ). Stimuli based on **wh-questions** (Fig.4) were identified as intended 72.92% of the time, while 11.72% were identified as commands and 15.36% as statements. The mean certainty rating was 3.72 and it was significant for wh-questions ( $p < 0.0001$ ,  $F = 24.37$ ). Stimuli based on **polar questions** (Fig.5) were identified as questions 45.31% of the time, while 21.70% were identified as commands and 32.99% as statements. The mean certainty rating was 3.70 and it was significant for polar questions ( $p < 0.0001$ ,  $F = 35.80$ ). According to the chi-square test, the observed value of 'intended' is related to its value for 'identified' at the 99% confidence level ( $p < 0.01$ ). Concerning the factor of certainty in all cases, the relevant chi-square test showed that it is related to its value for 'answer' ( $p < 0.01$ ). Sentence complexity is also proved to be a significant factor for sentence-type perception ( $p < 0.01$ ). An interesting observation is that simple sentences had lower rate of accurate identification (42.86%), compared to complex ones (50.82%). Furthermore, conjunction variation seems to influence listeners' choice ( $p < 0.01$ ).

## Discussion

The results of the present study indicate that sentence types in Greek may be identified through prosody, but it seems that this might not be the only distinctive cue for perception, since the identification rates were not very high in all cases. However, identification rates could not be expected to rise higher, because of the nature of the stimuli, which were obtained by synthetic modification (hum), as also proved by other studies with synthetic stimuli (e.g. Makarova, 2001). Thus, in the present experiment, an identification of 50-60% might be considered as an indicator of categorization between the 3 sentence types (statement, question, command). Apart from that, statistical analysis revealed that the listeners' choices were not incidental, since they were closely related to all contributing factors. **Wh-questions** are clearly 'marked' by their distinct prosodic structure, since there was very good correspondence between production and perception (72.92%). Acoustically, this may be attributed to the salient tonal rise-fall aligned with the wh-word and to the abruptly rising boundary tone, while the rest of the melody forms a low plateau (see Fig. 1). On the other extreme, **commands** seem to be quite difficult to perceive on the basis of prosody alone, since they had the lowest identification rates (35.42%) and they were equally confused with all the other sentence types. Their prosodic structure may vary quite a lot, mostly due to phrasing, while it generally resembles that of statements, especially at the falling boundary tone. **Polar questions** did not fall to very low identification rates (45.1%); the fact that they were confused with all sentence types may be acoustically attributed to their boundary tone, which forms a rise-fall tonal movement and might be

misleading for listeners. **Statements** had fairly good identification (52.08%) and they were mostly confused with commands, probably due to their low boundary tone. These findings agree with experiments in other languages which show that low or falling boundary tones elicit ‘declarative judgments’, and high or rising tones lead to ‘interrogative’ (e.g. Makarova 2001, Thorsen, 1980). Moreover, although tonal cues for sentence types are in general distributed throughout an utterance, the last stressed syllable and the poststressed ones are sufficient indicators for perception (Thorsen 1980).

Each sentence type has an acoustically distinct tonal structure, characterized in Greek by the type and location of tonal prominence (nucleus) and boundary tone (Baltazani, 2002; Botinis et al., 2000; Chaida, 2005, 2007). However, the findings of the present study with hum sound stimuli differ from our previous findings with natural stimuli (Chaida, 2005), which achieved 1:1 identification. Concluding, the difficulty that listeners had in identifying sentence types using only prosodic cues raises the question: is prosody the most significant factor for the perception of sentence types, or are other linguistic cues also needed?

### Acknowledgements

Many thanks to Charalabos Themistocleous for programming the perception tests, and to all the informants. I am grateful to Antonis Botinis for his guidance and to Marios Fourakis for his comments.

### References

- Baltazani, M. 2007. Intonation of polar questions and the location of nuclear stress in Greek. In: Gussenhoven, C. and Riad, T. (eds.), *Tones and Tunes*, Vol.II, *Experimental Studies in Word and Sentence Prosody*. Berlin: Mouton de Gruyter, 387-405.
- Botinis, A. 1998. Intonation in Greek. In: Hirst, D. and Di Cristo, A. (eds), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: CUP, 288-310.
- Botinis, A., Bannert, R. and Tatham, M. 2000. Contrastive tonal analysis of focus perception in Greek and Swedish. In: Botinis, A. (ed), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Kluwer Academic Publishers, 97-116.
- Chaida, A. (To appear). Tonal structures of complex sentences in Greek. *Proc. of the 8<sup>th</sup> International Conference on Greek Linguistics*. Ioannina, Greece.
- Chaida, A. 2005. Intonation of sentence types and focus in Greek. MA Thesis, University of Skövde, Sweden & University of Athens, Greece.
- Makarova, V. 2001. Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics* 29, 137-154.
- ’t Hart, J. 1998. Intonation in Dutch. In: Hirst, D. and Di Cristo, A. (eds), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: CUP, 96-111.
- Thorsen N. 1980. A study of the perception of sentence intonation - evidence from Danish. *Journal of the Acoustical Society of America* 67, 1014-1030.

# **Classification by discriminant analysis of the energy in view of the detection of accentuated syllable in Standard Arabic**

Amina Chentir<sup>1</sup>, Mhania Guerti<sup>2</sup> and Daniel Hirst<sup>3</sup>

<sup>1</sup>University of Blida, Algeria

<sup>2</sup>National Polytechnic School (ENP) of Algiers, Algeria

<sup>3</sup>UMR 6057, Laboratoire Parole et Langage (LPL), Aix-en-Provence, France

## **Abstract**

In this work, we exploited the acoustic parameter energy to a classification by a discriminant analysis to detect the primary accent syllables of type [CV] where [C] is a consonant and [V] a vowel, in Standard Arabic words. Generally, the three acoustic parameters used in prosody are: the fundamental, the duration and the energy, as well as the localization of the prominent syllable in final or initial position have an influence on the perception of the lexical accent in Arabic. A percentage of detection equal to 78% of the accentuated syllable shows the efficiency of such an approach that will be able to come to reinforce the methods based on the criteria of the fundamental, in the goal to improve the existing systems of synthesis and recognition in Standard Arabic.

Key words: classification by discriminant analysis, lexical accent, Standard Arabic, energy, accentuated syllable

## **Introduction**

Arabic is a Semitic language, it is classified among the oldest languages in the world. The Standard Arabic counts 34 phonemes: 6 vowels and 28 consonants. We will speak of the language Arabic in reference to what is called "the Unified Modern Arabic" or "the Standard Arabic", that is the language taught in the schools, written and spoken in the official context.

From the acoustic point of view, the prosody indicates the phenomena bound to the variation in the time of the parameters of fundamental frequency ( $F_0$ ), the intensity (I) and duration of emission (D) (Farinas 1998).

The accentuation is the setting in relief of a syllable in relation to those that surround it, and that are qualified therefore of not accentuated. These last years, we noted many works concerning the study of the Arabic prosody. What strengthened the hypothesis according to which the lexical accent exists in Arabic. Bohas (1979) admits the existence of the accent by showing that it plays a distinctive role. Rajouani (1989) confirmed that the detection of the primary accent seems sufficient for the study of the Arabic intonation and found from his experiments, the following result: the hierarchy ( $F_0$ , I, D) for the Arabic language.

### Arabic language and accent

For Ghalib, the accent exists in Arab but has no linguistic function and its importance is much lesser compared with English or Germany where it has incidences on the sense and the grammatical function of some words of the lexicon. Any isolated word in Arabic receives an accent which will be carried on the stressed syllable.

Al-Ani (1970) establishes the more used rules governing the place of the accent. He speaks about of the presence of three levels of accent: a first level or Primary Accent; a second level or Secondary Accent and a third level or Weak Accent. The position and the distribution of the accent depend on the number and the types of syllables contained in the word.

### Simulations and results

We have 4 Arabic-speaking: 2 men and 2 Women, who pronounce each one Arabic words having the following three-syllabic structure  $[S_1 S_2 S_3]$ :  $[C_1V C_2V C_3V]$  with  $[C_1]$ ,  $[C_2]$  and  $[C_3]$ , 3 different Arabic consonants and  $[V]$  a vowel. These words are pronounced inside carrier sentences. These sentences with  $[C_1V]$ , always corresponding to a syllable whose accent is primary. The recording was made in an anechoic room at the level of the Laboratoire Parole et Langage (LPL) then treated by a computer program PRAAT which can analyse, synthesize and manipulate speech data.

In our approach, we followed the following stages:

Stage 1: Segmentation and phonetic transcription of the recorded words

Stage 2: Extraction then the calculation of the medium-term spectre for every vowel detected inside the used word

Stage 3: Make a discriminant analysis to classify all the vowels in an orderly structure and create the appropriate configuration

Stage 4: Generate the matrix of the confusions to verify the conformity of the predictive classification with the reality

Stage 5: Consider values for additions of vowels not being present in the sample of training. We shall so manage to predict the values of new observations in the classification or the already existing regrouping

Stage 6: Generate the corresponding matrix of confusion.

To be able to interpret the results, we exploited the method of the bootstrap (Efron and al. 1994) in our corpus. This method gets organized around a technique of re-sampling, accompanied with a “big” number of iterations which result from the application of the method of Monte Carlo (Landau and al. 2000).

We proceeded to the learning of sentences of our corpus. Once realized, we passed to the recognition of other sentences not included in the first phase of calculation. We proceeded to the calculation of the matrix of total



confusion for every sentence as well as the percentage of correct affectations.

To end on the efficiency of the used method, we appealed to the principle of the method of Bootstrap (defined previously) and we then calculated the matrix of total confusion corresponding to the tested corpus.

Table 1. Matrix of confusion in learning (L) and in recognition (R) and the percentages of affectation obtained according to every sentence.

Removed sentence Speakers $X_1 - X_2$	Sentences		
$X_1 - X_2$	540L	424	87 29
		92	320 128
		33	241 266
	60R	62.35 %	
		S1 : 78.52 %	
		S2 : 59.26 %	
$X_1 - X_2$	540L	47	10 3
		12	33 15
		6	27 27
	60R	59.44 %	
		S1 : 78.33 %	
		S2 : 55 %	
$X_1 - X_2$	540L	33	241 266
		92	320 128
		424	87 29
	60R	49.26 %	
		S1 : 78.33 %	
		S2 : 55 %	

We obtained then the Table 1 which allows us to conclude as follows: the phase of learning gives a good percentage of recognition equal to 62.35%. It is clear that it is the classification of both unaccented syllables ( $S_2$  and  $S_3$ ) that are at the origin of this reduction, the accentuated syllable  $S_1$  is classified with a good rate equal to 78.52% and the global phase of test is very slightly superior to the threshold corresponding to an unpredictable classification. However, we note the very good classification of the syllable  $S_1$  (78.33%).

## Conclusion

In this work, we took advantage of the classification by discriminant analysis based on the acoustic parameter energy to detect the primary accent in SA in the syllable of type [CV]. Our choice limited itself to the three-syllabic Arabic words. After having segmented and transcribed manually the used corpus, we applied him our algorithm based on the discriminant analysis. A percentage of detection equal to 78% of the accentuated syllable has been obtained.

This is only a first approach in the detection of the primary accent in standard Arabic by a discriminant analysis of the prosodic parameter energy. The obtained results are to be tested on more important corpuses of Arabic. But already, we can say that the classification by discriminant analysis of the



criterion energy can be a supplementary parameter for the detection of accentuated syllables what can come to enrich the methods of recognition already existing, based only on the criterion of the fundamental.

### References

- Al-ani S.H. 1970. Arabic phonology: An acoustical and physiological investigation, Mouton and Co. (Ed.), The Hague, Netherland.
- Bohas G. 1979. Contribution à l'étude de la méthode des grammairiens arabes en morphologie et en phonologie d'après les grammairiens arabes tardifs. Thèse de doctorat, Université de Lille 3, France.
- Efron B., Tibshirani R.J. 1994. An Introduction to the Bootstrap, Chapman & Hall/CRC (Ed.), USA.
- Farinas, J. 1998. La Prosodie pour l'Identification Automatique des Langues, Rapport de DEA, Université Paul Sabatier, France.
- Ghalib M. (ND). Etude de quelques aspects de l'intonation en Arabe, Revue de l'Université de Bassorh, Vol. 10, pp. 198-228, Irak.
- Landau D.P., Binder K. 2000. A guide to Monte Carlo simulations in statistical physics, Lavoisier (Ed.), France.
- PRAAT, Praat : doing phonetics by computer, [www.praat.org](http://www.praat.org).
- Rajouani A. 1989. Contribution à la réalisation d'un système de synthèse à partir du texte pour l'arabe, thèse de doctorat, Université Mohamed V, RABAT, Maroc.

# **The identification of the place of articulation in coda stops as a function of the preceding vowel: a cross-linguistic study**

Man-ni Chu<sup>1,2</sup>, Carlos Gussenhoven<sup>2</sup> and Roeland van Hout<sup>2</sup>

<sup>1</sup>Department of Linguistics, National Tsing-Hua University, Taiwan

<sup>2</sup>Department of Linguistics, Radboud Nijmegen University, the Netherlands

## **Abstract**

Three groups of listeners, one with a /p,t,k/ stop system in the syllable coda (Dutch), one with /p,k,ʔ/ (ChaoShan) and one with /p,t,k,ʔ/ (ZhangQuan) were asked to identify the final C in CVC-stimuli as one of the consonants [p,t,k,ʔ]. Identification was affected by the nature of the coda C and the V. The identification of the bilabial stop significantly increases after /i/, that of the velar stop after /-a/ and that of the alveolar stop after /u/.

Keywords: place of articulation, coda, Dutch, ChaoShan, ZhangQuan

## **Introduction**

A number of studies (Delattre et al. 1955 and many) indicated that second formant transitions are crucial for coda consonant identification. In this study, we extend the research to three groups of subjects with different language backgrounds, ChaoShan with an unreleased [p,k,ʔ] coda system, Dutch with a released [p,t,k] system and ZhangQuan with an unreleased [p,t,k,ʔ] system.

## **Materials and Method**

We used CVC stimuli, in which the onset C was varied over six consonants (p, t, k, p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>), the V over three vowels (i u a), and the unreleased coda C over four (p t k ʔ), which were spoken by a male speaker of Tainan. We manipulated the f<sub>0</sub> so that each stimulus occurred in a low-toned version (123Hz) and a high-toned version (164Hz). The level of noise (0db, -10db, and -20db) were varied in the two Chinese languages, while Dutch had only the first two noise levels. Because of lexical gaps in Tainan, we ended up with 48 original sound files, instead of 6\*3\*4=72. There were therefore 48 \* 2 (tones) \* 3 (noise levels) = 288 stimuli for the Chinese languages and 192 for Dutch, which were presented auditorily in a different random order to each subject. 54 (Chaoshan), 16 (Dutch) and 37 (ZhangQuan) participants were asked to identify the coda as one of [p,t,k,ʔ] by pressing one of the four.

## Results

The cross-tabulations and the contingency coefficients indicated that the nature of the vowel and of the coda significantly affected subjects' identifications. Neither tone nor noise level had any effect on the responses. We applied two series of analyses, one with the overall coda response as the dependent variable, the other one on the correct responses per target coda. The repeated measures ANOVAs were done for the language groups separately, with vowel category as the independent variable. The vowel effect was tested by using the Huynh-Feldt adapted F values. The results of the post-hoc comparisons reported were adjusted by applying the Bonferroni procedure.

The overall coda response analyses showed that the coda /k/ was favored after /a/ in all three language groups ( $F(1.73, 91.61)=66.05$ ;  $F(1.54, 23.04)=56.61$ ;  $F(1.89, 67.16)=34.17$ ; all  $p<.05$ ), where the results are for the Chaoshan, Dutch and ZhangQuan listeners, respectively. The coda /p/ was favored when occurring after /i/ ( $F(1.91, 101.39)=93.18$ ;  $F(1.44, 21.65)=47.85$ ;  $F(1.9, 68.0)=47.13$ ; all  $p<.05$ ). The post-comparisons showed that the ZhangQuan listeners had a significant preference of /i/ and /a/ over /u/, but the difference between /i/ and /a/ was not significant ( $p=.09$ ). Coda /t/ was favored after /u/ ( $F(1.27, 67.31)=117.9$ ;  $F(1.71, 25.62)=41.32$ ;  $F(1.32, 47.58)=9.93$ ; all  $p<.05$ ). The post-comparisons showed that the ZhangQuan listeners had a significant preference of /u/ and /i/ over /a/ but the difference between /u/ and /i/ was not significant ( $p=1.0$ ). In addition, the /?/ was significantly least favored after /a/ in the Dutch group ( $F(1.86, 27.9)=16.96$ ,  $p<.05$ ), but it was favored by ZhangQuan listeners after /u/ ( $F(1.54, 55.37)=17.59$ ;  $p<.05$ ). The ChaoShan natives did not show any preference ( $F(1.7, 89.9)=1.41$ ,  $p>.05$ ). The conclusion for the /?/ is that there is no cross-linguistic pattern of preference. Figure 1 presents the mean responses for the four coda consonants as a function of preceding vowel for each language group.

The percentage correct analyses showed that the coda /t/ was favored when occurring after /u/ ( $F(1.36, 72.04)=114.33$ ;  $F(2, 30)=34.4$ ;  $F(1.56, 55.92)=5.36$ ; all  $p<.05$ ). The post-comparisons showed that the ZhangQuan listeners had a significant preference of /i/ and /u/ over /a/, but the difference between /i/ and /u/ was not significant ( $p=1.0$ ). The coda /p/ was favored when occurring after /i/ by the ChaoShan and Dutch groups ( $F(1, 53)=71.59$ ;  $F(1, 15)=7.79$ ; all  $p<.05$ ), while ZhangQuan listeners did not show any preference ( $F(1, 36)=2.8$ ,  $p>.05$ ). In addition, the /?/ after /a/ was significantly favored by the ChaoShan listeners ( $F(1.95, 103.45)=11.96$ ) but it was not favored by Dutch listeners ( $F(1.51, 22.66)=6.95$ ). The /?/ was favored after /u/ by the ZhangQuan listeners ( $F(1.95, 70.14)=6.43$ ). The post-comparisons showed that they had a significant preference of /u/ over /i/, but

/a/ was not significantly different from /u/ or /i/ ( $p=.25$  and  $p=.14$ ). Figure 2 presents the mean percentage correct responses for the three codas for each language group. Overall, /p/ was mostly favored after /i/, /k/ was favored after /a/, /t/ was favored after /u/, and there is no clear pattern for /?/. The stimuli ended in /k/ only occurred in the /ak/ condition, due to lexical gaps in Tainan.

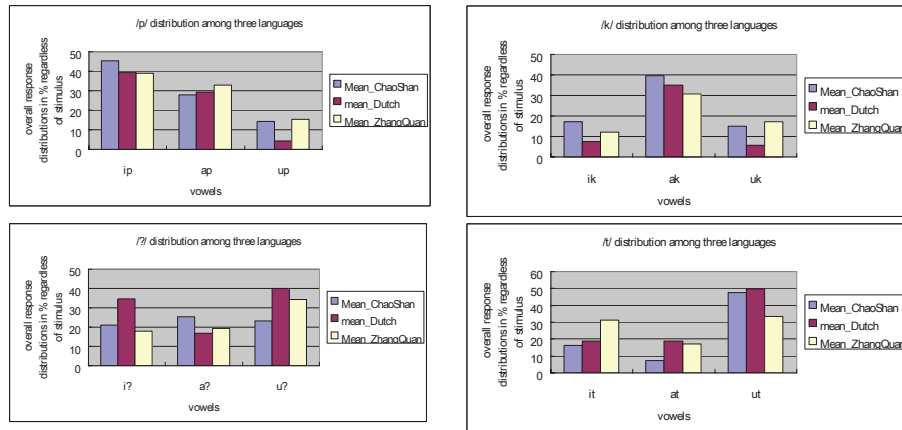


Figure 1: Identification of coda [p,k,?,t] as a function of the preceding vowel for ChaoShan, Dutch and ZhangQuan listeners.

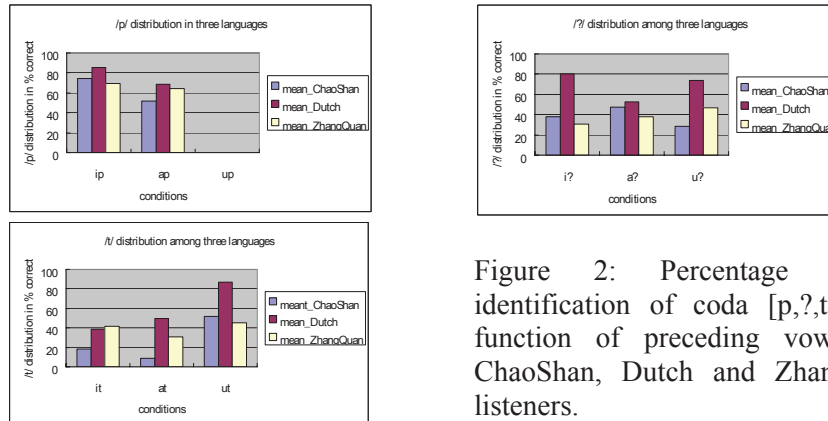


Figure 2: Percentage correct identification of coda [p,?,t] as a function of preceding vowel for ChaoShan, Dutch and ZhangQuan listeners.

## Discussion

Our results confirm those by Liberman et al. (1952) that coda stop identification varies with the spectral properties of the preceding vowel: /-p/ is preferred after /i/ and /-k/ after /-a/. The relatively high F2 of the high front vowel [i] causes a falling transition, which resembles the locus effect of

[-p], while the relatively low F2 of [a] causes a rising transition which resembles the locus effect of [-k]. In addition, the locus effect of [t] causes a rising transition when accompanied with the lowest F2 of the back vowel [u] (Delattre et al. 1955), which is a prominent feature compared to other consonants neighboring /u/. Formant transition and release bursts are two important cues for stop identification (Hall et al. 1956 and others). Since the identification of unreleased stops necessitates subjects to rely on formant transitions only, significant biases in the perception of final stops will be caused by vowel qualities that bring out those transitions. On the other hand, there is no consistent effect of the vowel in glottal stop identification, because there are no locus transitions.

Since the three language groups responded in very similar ways, we conclude that the biasing effects are language-independent. The absence of /t/ in ChaoShan, the absence of phonemic /ʔ/ in Dutch and the presence of the full four-way contrast in ZhangQuan had no interpretable effect on the perception of these consonants by the three groups of listeners.

### **Acknowledgements**

We would like to thank Joop Kerkhoff and all the participants.

### **References**

- Delattre, P.C., Liberman, A.M., Cooper, F.S., 1955. Acoustic Loci and Transitional Cues for Consonants. *The Journal of the Acoustical Society of America* 27(4), 769 -773.
- Halle, M., Hughes, G. W., Radley, J.-P. A., 1957. Acoustic properties of stop consonants, *The Journal of the Acoustical Society of America* 29, 107-116.
- Liberman, A.M., Delattre, P.C., Cooper, F.S., 1952. The Role of Selected Stimulus Variables in the Perception of the Unvoiced-stop Consonants. *The Journal of the Acoustical Society of America* 65, 497-516.

## **Reading mathematical exercises: preliminary results**

Deolinda Correia, Isabel Hub Faria and Paula Luegi

Laboratório de Psicolinguística, LinSe, Centro de Linguística da Universidade de Lisboa

### **Abstract**

This research on reading comprehension is concerned with linguistic complexity processing and solving of mathematical exercises by 95 Portuguese students aged 9 to 15 years, attending one of the three basic school levels. Exercises were selected from the full set of national examinations (from 2000 till 2007), covering different mathematical areas. Experimental task was to present subjects with an exercise text followed by a possible result which they had to evaluate as a good or bad answer. Preliminary results indicate that although the extension of the exercise text and of the answer influences the time spent reading and solving the problem, it does not necessarily make the resolution of the exercise harder.

Key words: mathematical exercises processing/solving and reading

### **Introduction**

The present study is part of a larger research project on reading comprehension of words, sentences and texts that is taking place at the University of Lisbon.

This paper particularly focuses on mathematical exercises resolution by subjects attending the three basic school levels.

The Portuguese Ministry of Education (ME) has pointed out as one of the major reasons for the lack of success in mathematics evaluation the difficulties of basic school students in reading and comprehending the examinations' texts. However, these declarations were never rigorously justified, since until the present, no studies revealed the reading behaviours of the subjects while processing mathematical exercises. Studies in this area (Hegarty, Mayer and Green 1992; Thompson 1992; Suppes 1990) point out mainly to factors associated with mathematical competence.

We selected from the whole set of mathematics national examinations the set of stimuli used in our experiment.

The structural diversity under observation maintains the thematic domains and the types of operations of the official ME program in mathematics for each basic level.

In this study we analysed the 'yes/no' answers provided by 95 students to 'true/false' results of 60 exercises. Each result was shown on the screen after the participant had read the exercise text.

## Experiment

### Methodology

#### Material

Exercises used as experimental stimuli have been selected from the full set of national examinations (from 2000 till 2007).

Exercises cover different areas which are defined by the national mathematical educational programs: *numbers and operations, shapes and space, measures and dimensions* (4<sup>th</sup> year); *numbers and calculus, statistics and probabilities, proportionality* (6<sup>th</sup> year); *numbers and calculus, statistics and probabilities, algebra, geometry* (9<sup>th</sup> year), and require the following aspects of mathematical competence: concepts and procedures, reasoning, problems solving and communication. The nature of the exercises used as stimuli was unimodal (only text) and bimodal (text and image).

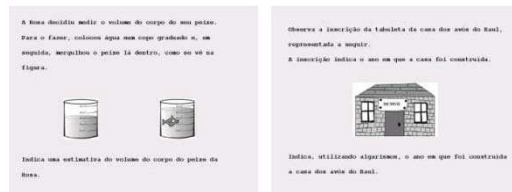


Figure 1. Examples of the exercises, from the examination test, presented.



Figure 2. Examples of the answers provided.

The number of presented stimulus was different through the levels: 1<sup>st</sup> level – 24 exercises; 2<sup>nd</sup> level – 16 exercises; and 3<sup>rd</sup> level – 20 exercises.

### Method

Each exercise (Figure 1) was presented on a computer screen followed by the respective possible answer (Figure 2). Participants had to read the exercise text and press a button to answer the question. After reading the provided answer, participants had to press a certain button if the answer was true or another button if the answer was false.

Time spent reading the exercise text and the answer was registered with E-Prime. The accuracy of the answer was also registered.

## Participants

The sample is constituted by 95 subjects, male and female, aged 9 to 15 years, attending the last levels of the basic cycles in national Portuguese public schools. All the participants were Portuguese native speakers and made their entire schooling in Portugal. Subjects were distributed according to three different levels: 1<sup>st</sup> level – 32; 2<sup>nd</sup> level – 33; and 3<sup>rd</sup> level – 30.

## Results

We classified all the exercises for the following variables: Thematic Domain, Mathematical Operation Type, Number of words on the exercise text, Number of words on the answer text, Number of images on the exercise text, and Number of images on the answer text. These variables were contrasted to the participants' behaviour: Time spent on the exercise text, Time spent on the answer text, and Number of right answers.

First cycle results show that the higher number of subjects' right answers (91%) occurs in *numbers recognition*, specifically *roman numbers reading* and the area with worst results (38%) was *estimative calculation* in the Thematic Domain of *measures and dimensions*. On average, the accuracy of answer is above 50%. In this cycle, there are statistically significant positive correlations between the time spent on the exercise text and the number of words ( $p < 0,05$ ), and between the time spent on the answer exercise and its number of words ( $p < 0,01$ ). The same contrast was also found in the third cycle, but only as far as the exercise text is concerned ( $p < 0,01$ ). In the third cycle, we find in the same Thematic Domain the best and the worst results. For instance, in *statistics and probabilities*, participants had 93% in one exercise and 17% in another.

No correlation of the extension of the text and reading time was found on the second cycle. However, there is a positive correlation between the time spent on the answer text and its number of images, i.e., the higher the number of images in the answer, the higher the time spent on it. More interestingly, there is also a positive correlation between the number of right answers and the number of words of the answer text. This means that the provided answers that have an extended composition are associated with a very high score of subjects' right answers.

In this cycle, we concluded that for the same Mathematical Operation Type participants had different results depending on the Thematic Domain, for instance they had high scores in *number and calculus* and low scores in *proportionalities*.



## Discussion

These preliminary results indicate that although the extension of the exercise text and of the answer influences the time spent on reading and solving the problem, it does not necessarily make the resolution of the exercise harder. In other words, the difficulties in mathematical problems solving, and, consequently, the high level of insucces in this area of knowledge, does not seem to be due to the extension (at least in number of words) of the exercises.

The Mathematical Operation Type and the Thematic Domains are also determinant for the exercise resolution. In all cycles, but with more evidence in the first one, *numbers and operations* is always the domain with the highest number of right answers.

In face of these results, we will look, in further analyses, into correlations between linguistic complexity of the exercises texts (not always very well controlled) and the dependent variables (time spent on exercise text and on answer text, and accuracy of answer) analysed in this experiment. Besides, we will also contrast the eye movement's data collected while solving the exercises with the exercise characteristics and with the linguistic complexity of the exercises.

## References

- Hegarty, M., Mayer, R.E. and Green, C. 1992. Comprehension of arithmetic word problems: evidence from students' eye fixations. *Journal of Educational Psychology*, 84, 76-84.
- Hoffman, J. E. 1998. Visual attention and eye movements. In H. Pashler (Ed.) *Attention*. London: University College London Press.
- Kiess, H. and Bloomquist, D. 1985. *Psychological research methods: a conceptual approach*. Boston: Allyn and Bacon.
- Just, M. A. and Carpenter, P. A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Suppes, P. 1990. Eye-movement models for arithmetic and reading performance. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and Cognitive Processes*, 455-477. New York: Elsevier Science publishing.
- Thompson, P. 1992. Notations, conventions and constraints: contributions to effective uses of concrete materials in elementary mathematics. *Journal for Research in Mathematics Education*, 23, 123-147.

# Prosody in read aloud text: relation with information status, content type and boundary strength

Hanny den Ouden<sup>1</sup> and Carel van Wijk<sup>2</sup>

<sup>1</sup>Faculty of Humanities, UiL/OTS, Utrecht University, The Netherlands

<sup>2</sup>Faculty of Humanities, Tilburg University, The Netherlands

## Abstract

In two short stories three text characteristics have been implemented independent of each other: information status, content type, and boundary strength. Thus, each characteristic could be assessed for its unique effect on prosody. Both stories were read aloud by ten speakers. No effects were found for pitch range and articulation rate, but pause duration did vary systematically with each of the text characteristics. Effects ranged from small for information status to moderate for content type, and large for boundary strength.

Key words: narrative, text characteristics, prosody, pause duration

## Introduction

Research on text prosody has focused on various prosodic features of sentences in relation with their position and function in a text (see e.g. den Ouden, Noordman and Terken, 2008). It has not settled the issue, however, whether effects are real or spurious, that is, does a specific text characteristic exert an influence on its own, or because it happens to coincide with another characteristic actually being the effective one? In this paper we report a study in which, independent of each other, three text characteristics were assessed for their effect on prosody.

## Text characteristics

Three text characteristics have been shown to affect prosody: information structure (van Donzel, 1999), message content (Swerts, Krahmer, Theune and Weegels, 2002), and text organization (den Ouden, 2004). Although some use these terms interchangeably, they represent different perspectives on a text as we show here with the texts read aloud in the study.

Information structure is signaled by referential continuity. Does a sentence give information on a referent mentioned in the sentence directly preceding or on some other referent? Box 1 shows how successive sentences switch between both possibilities, denoted [*same*] and [*other*].

Message content is defined as the global meaning or gist of a paragraph. Stories are constructed from four basic content types (see Box 2). They start with some background information (*state*) that leads to a challenging situation (*outcome*). This takes story characters into a sequence of episodes

in which they want things (*goal*), try to realize them (*attempt*), and experience that they have either failed or succeeded (*outcome*).

Box 1. Illustration of referential continuity (either [same] or [other]).

..... (11) Recently, a businessman had a promising contact with a Colombian (12) He [same] could buy pure cocaine at a low price (13) The police [other] wanted to take advantage of this contact (14) They [same] decided to place a trap for the mafia (15) They [same] wanted to strike drugs traffic very hard (16) Crime in the street [other] had to be pushed back (17) The police [other] formed a team of detectives (18) That [same] had a number of officers infiltrate .....

Box 2. Illustration of content types (for each instance the serial number and content of the central sentence, and serial numbers of surrounding sentences).

State	(6)	Circle Island has a shortage of water	1-10
outcome	(11)	Recently, a scientist discovered a cheap method of water treatment	11-12
goal	(14)	The farmers decided to build a canal straight across the island	13-16
attempt	(18)	They lobbied a few senators to join	17-21
outcome	(23)	The senate, however, voted against	22-24
goal	(26)	A smaller canal had to be build	25-27
attempt	(28)	The construction of the smaller canal started	28-30
outcome	(33)	The canal project had failed	31-34

Text organization is scored as boundary strength: the distance between adjacent sentences in terms of the hierarchical structure of the text like the one depicted in Figure 1. Boundary strength is *absent* for short distances (e.g., from 9 to 10), *weak* and *moderate* for longer distances (e.g., from 4 to 5 and 12 to 13), and *strong* for long distances (e.g., from 16 to 17; for details on scoring, see den Ouden, 2004, p.25).

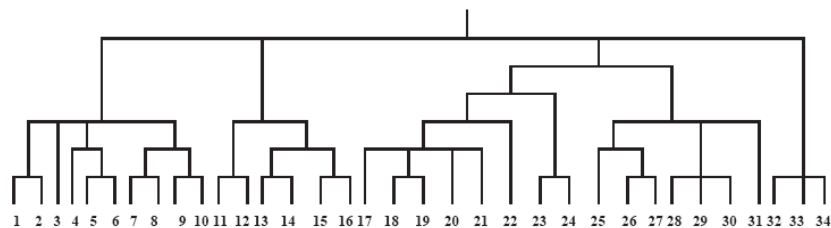


Figure 1. Structure underlying the experimental texts.

## Research questions

Does each of the three text characteristics affect prosody independent of the other two? And if so, how much do they differ in their effect sizes?

## Method

### Material

Experimental texts were Thorndyke's (1977) *Island* classic on a political issue and a newly written story on drug crime. Both texts were 34 sentences long. They differed in their contents, but shared the hierarchical structure presented in Figure 1. Each sentence was scored for Information Status (it started with a referent mentioned in directly preceding sentence, or it did not), Content Type (it was part of a paragraph informing on a state, goal, attempt, or outcome), and Boundary strength (its distance from the directly preceding sentence in the hierarchical structure was scored as either absent, weak, moderate, or strong). Within a story each of the text characteristics was distributed sufficiently even, both in number and position. They occurred independent of each other with one exception: strong boundaries tended to coincide less often with a 'same' referent. This trend was too weak to be of any influence on the results.

### Procedure

Ten speakers participated (5 men, 5 women). They prepared both texts and then read them aloud (with a third structurally different text read aloud in between the two). Three prosodic features were measured of each sentence: duration of preceding pause, pitch peak and articulation rate. For each prosodic feature repeated measures anova's were run with Gender and Speaker as nested within-factors, with Story and one of the three text characteristics as between-factors, and with the other two text characteristics as covariates. Effect size was determined with partial eta-squared ( $\eta^2$ ).

## Results

Pitch peak and articulation rate showed no significant main effects, but pause duration did: for Information Status ( $F(1,60)=3.27$ ,  $p=.07$ , i.e., tested one-sided  $p<.05$ ,  $\eta^2=.05$ ), Content Type ( $F(3,56)=3.06$ ,  $p<.05$ ,  $\eta^2=.14$ ), and Boundary Strength ( $F(3,56)=10.13$ ,  $p<.001$ ,  $\eta^2=.35$ ). Table 1 presents mean pause durations for the three text characteristics. All effects occurred independent of Gender and Story; there were no significant interactions. The effects on pause duration differed in size: small for Information Status, medium for Content Type, and large for Boundary Strength. To illustrate the magnitude of the differences Table 1 also presents the difference of each score from the overall mean pause duration (718 msec).

Table 1. Pause duration in relation with text characteristics (first line: score in absolute figures, second line: score relative to overall average of 718 msec).

Information Status		ContentType				Boundary Strength			
same	other	state	outcome	goal	attempt	absent	weak	moderate	strong
685	746	653	694	767	775	660	696	738	948
-33	+28	-65	-24	+49	+57	-58	-22	+20	+230

## Discussion

When carefully controlled for confounding effects in the design both of the experimental texts and the statistical analyses, each of the three text characteristics, that is, information status, content type and boundary strength, appears to have a discernable effect on pause duration independent of the other ones. Linguistic theorizing on text prosody and computational generation of artificial speech will have to consider these three text characteristics simultaneously to gain a full understanding of human speech and an adequate mimic in text-to-speech systems.

## References

- Donzel, M. van 1999. Prosodic aspects of information structure in discourse. Doctoral dissertation, University of Amsterdam, The Netherlands.
- Ouden, H. den 2004. Prosodic realizations of text structure. Doctoral dissertation, Tilburg University, The Netherlands.
- Ouden, H. den, Noordman, L. and Terken, J. 2008. Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication* (in press).
- Swerts, M., Krahmer, E., Theune, M. and Weegels, M. 2002. The dual of denial: Two uses of disconfirmations in dialogue and their prosodic correlates. *Speech Communication* 36, 133-145.
- Thorndyke, P. 1977. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9, 77-110.

# **Imperatives in European Portuguese: a perception approach**

Isabel Falé

Laboratório de Psicolinguística, LinSe, CLUL, Portugal, and  
Universidade Aberta, Portugal

## **Abstract**

In European Portuguese (EP), intonation has a grammatical function. In the available research in EP intonation, production imperatives are said to have an F0 rising-falling contour of large amplitude and are described as having an intonation contour similar to wh-questions, although they present rhythmic differences (Viana, 1987). A possible increase in pitch span was also reported. Two perception experiments were developed to identify sound-sequence features of imperative intonation prototype. The results showed that the major intonation distinction between declaratives and imperatives in EP was related both with local events, that determine utterance contour shape, namely F0 movements, and global events that locate F0 levels.

Key words: speech prosody, intonation, imperatives, speech perception

## **Introduction**

The available research on European Portuguese (EP) intonation has not so far systematically addressed the issue of imperatives. Imperatives were said to have an F0 rising-falling contour of large amplitude and were described as having an intonation contour similar to wh-questions, although they presented rhythmic differences (Mateus et al. 1983, Viana 1987). A possible increase in pitch span was also reported.

Imperative sentences are syntactically and morphologically marked in EP. In general, these grammatical features may be sufficient to distinguish imperatives from other sentence types. However, the intonation features of imperatives seem to be quite prominent and play an important role in EP perception and processing.

Two perception experiments were developed to identify sound-sequence features of imperative intonation prototype. All sound sequences were recorded by two EP native speakers: a female (NA) and a male (LL).

## **Experiment I**

Experiment material was composed by 195 sound sequences (sentences) of different sentence types: declaratives, questions, wh-questions, imperatives and exclamations. Several variables such as segmental constituency, primary stress location, syllable structure, sentence syntactic constituency and

illocutionary strength were controlled. Experiment I task consisted in listening to sound sequences and immediately categorizing them in four sentence types previously defined. This task recruited *top-down* linguistic data processing and linguistic explicit knowledge. 40 EP native speakers, aged between 19 and 50, with no history of hearing or language deficits or disorders, participated in the experiments.

## Results I

The inclusion criterion of a sound sequence in one of the available categories was a recognition result equal or higher than 75%.

85.3% of the 35 imperatives present in the *corpus* were categorized as imperatives. An acoustic and phonetic analysis using *Praat* software was performed on these sequences. Based on earlier studies on EP (Viana 1987, Frota 1998, Mata 1999), all of the sentences were labelled according to specific phonetic points believed to be the most informative ones for intonation analysis: onset of the sentence (O); first stressed vowel (FSV); F0 peak (FP); final pre-stressed vowel (FPSV); last stressed vowel (LSV); last vowel or voiced consonant (LVC) and data were collected.

For local events analysis, all F0 movements were categorized according to their direction (rising, falling, rising-falling, falling-rising, flattened), movement amplitude and segmental alignment. Tonal events alignment with segmental structure in EP is directly related with stressed vowel/syllable location (Frota 1998, Mata 1999, Grønnum and Viana 1999). For global intonation events study, pitch register (Patterson and Ladd 1999) pitch level (Rietveld and Vermillion 2003) and pitch span were considered.

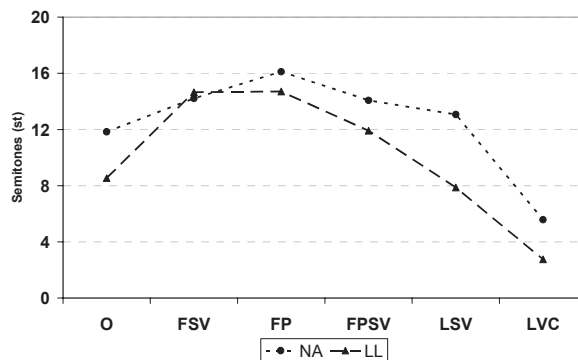


Figure 1. Representation of F0 average topline values in imperative sentences produced by two speakers (NA) and (LL), in semitones.

The analysed imperatives contour shape showed an initial rise from the onset (O) to the F0 peak (FP). In these sentences, FSV occurs in the verb. After FP, begins a falling movement of large amplitude, that is more visible from the final pre-stressed vowel (FPSV) or the last stressed vowel (LSV) to the last vowel or voiced consonant (LVC) (see Figure 1).

The major intonation distinction between declaratives and imperatives is related both with local events, that determine utterance contour shape, namely F0 movements, and global events that locate F0 levels (see Tables 1 and 2). A Principal Component Analysis of imperatives data revealed that variance in imperative sentences is due to FSV, FP, LSV, FPSV, Pitch Register and Pitch Span variables. This result was also corroborated by a Classificatory Analysis that joined these variables in the same cluster.

Table 1 and Table 2. Average F0 values and respective standard deviations of Pitch Level, Pitch Register and Pitch Span in imperative (*Imper.*) and declarative sentences (*Declar.*), produced by the female speaker (NA) and the male speaker (LL), in semitones.

NA	Pitch Level		Pitch Register		Pitch Span	
	A	SD	A	SD	A	SD
Imper.	3	1,3	12	2,3	14	2,8
Declar.	3	1,0	9	0,6	10	1,7
LL	Pitch Level		Pitch Register		Pitch Span	
	A	SD	A	SD	A	SD
Imper.	2	0,32	11	2,68	13	2,62
Declar.	2	0,27	7	1,02	9	1,73

## Experiment II

A categorical perception experiment with *hummed* sentences was developed. From two natural sentences, one produced by a male speaker and another by a female, two multi-step *continua* from each sentence were created, from declarative to imperative contour, through acoustic manipulation (PSOLA) and submitted to 20 EP native speakers that performed two tasks: an identification and a discrimination task. For the identification test, subjects had to categorize each presented stimulus either as a declarative or as an imperative sentence. Each of the 14 stimuli was repeated eight times in random order. In addition to response data, reaction times were also collected. For the discrimination task subjects had to decide whether the stimuli in each pair (13 pairs repeated five times) were equal or different. Experimental design and procedures were developed with *E-Prime*.



## Results II

Identification test results revealed a uniform evolution that is dependent on the F0 value of the Last Stressed Vowel: sentences that were classified as imperatives presented high F0 values in this variable. When the value of this variable decreases, the identification as imperative also diminishes. However, general results of this test were not very clear as far as imperative intonation category is concerned. The use of *hummed* sentences may have contributed to this situation.

## Conclusions

The prototype of imperative category in EP is clearly related to high F0 global values and to a particular intonation shape described earlier. Acoustic and phonetic analyses of both perception experiments point to the high probability of the Final Pre-Stressed Vowel and the Last Stressed Vowel variables being the most informative and prototypical to define imperative sentence category in EP.

## References

- Falé, I. 2005. Percepção e Reconhecimento da informação entoacional em Português Europeu. PhD Dissertation, University of Lisbon.
- Falé, I. and Faria, I. H. 2007. Imperatives, orders and requests in European portuguese intonation. Proceedings of the 16th International Congress of Phonetic Sciences, 1041-1044.
- Frota, S. 1998. Prosody and focus in european portuguese. PhD Dissertation. University of Lisbon.
- Mata, A. I. 1999. Para o estudo da entoação em fala espontânea e preparada no Português Europeu. Metodologia, resultados e implicações didáticas. PhD Dissertation. University of Lisbon.
- Mateus, M.H., Brito, A.M., Duarte, I. and Faria, I.H. 1983. Gramática da língua portuguesa. Lisboa: Editorial Caminho.
- Patterson, D. and Ladd, R. 1999. Pitch range modelling: linguistic dimensions of variation. XIV ICPhS Proc. San Francisco, 1169-1172.
- Rietveld, T., Vermillion, P. 2003. Cues for perceived pitch register. *Phonetica*, 60, 261-272
- Viana, M.C. 1987. Para a síntese da entoação do português. PhD Dissertation, University of Lisbon.

## Nasometric values for European Portuguese: preliminary results

Isabel Falé<sup>1,3</sup> and Isabel Hub Faria<sup>1,2</sup>

<sup>1</sup>Laboratório de Psicolinguística, LinSe, CLUL, Portugal

<sup>2</sup>DLGR, Faculdade de Letras, Universidade de Lisboa, Portugal

<sup>3</sup>Universidade Aberta, Portugal

### Abstract

Nasal sounds frequencies in European Portuguese represent 21% of Português Fundamental *corpus* sounds (Nascimento et al. 1987), revealing how nasality plays an important role in this language and how a speech problem affecting nasality can interfere severely in one's speech intelligibility. In order to obtain the first standard nasometric values for European Portuguese we developed two tests (syllable repetition and text reading) and collected data from 25 adults. Preliminary results showed that: oral stimuli achieved an average nasalance score of 10%; syllables with nasal consonant and nasal vowel achieved 77% and the reading passages with nasal saturation presented an average score of 44%. Considering these results, we acknowledge the existence of three different levels of normal nasality.

Key words: nasometry, nasality, speech production, speech assessment

### Introduction

Resonance is one of the most prominent characteristics of human speech. Problems affecting nasal resonance are widely reported in speech pathologies, causing severe perturbations in subjects speech intelligibility.

Nasal resonance values can be obtained through different instrumental methods, however, the nasometer from Kay Elemetrics 6200-3 proved to be an indirect and objective assessment instrument which results have shown to have a high correlation with perceived nasality (Hardin et al. 1992, Hirschberg et al. 2005). This nasometer provides a nasalance score which corresponds to a ratio of nasal to nasal-plus oral acoustic energy, multiplied by 100.

Normative nasalance scores have been already determined for several languages such as English (Seaver et al. 1991), French, Spanish (Nichols 1999), Puerto Rican Spanish (Anderson 1996), Finnish (Haapanen 1991), Dutch, Flemish (van Lierde et al. 2000) and Hungarian (Hirschberg et al. 2005). Some of the studies pointed to the existence of differences among different dialects, reinforcing the need to have nasalance norms established for each sample in every language.

Our research main goal was to collect normative nasalance scores for European Portuguese language, in order to make possible an easy detection

of nasal resonance problems in this language, either hiponasality or hipernasality.

### **Experiment**

In European Portuguese there are three nasal consonants and five nasal vowels. Usually, nasal consonants spread their nasal feature to the adjacent vowels nasalizing them. Considering these language specificities, we developed two tests: a repetition syllable test and a text reading test.

### **Experimental material**

Starting from one sub-test of the Mackay-Kummer Test (SNAP) - repetition syllable test - we built a new one, the TN-PE, which raised the number of pre-existent syllables in order to assess language specific consonant-vowel combinations, namely those with nasal vowels. Forty syllables distributed by four combinations of resonance type syllables were considered: A. oral consonant – oral vowel (13 syllables); B. oral consonant – nasal vowel (8 syllables); C. nasal consonant – oral vowel (12); D. nasal consonant – nasal vowel (7).

We also created eight reading passages considering two basic different criteria: absence of nasal sounds and saturation of nasal sounds (26 to 33% of nasal sounds in each). Neither of them was balanced in respect to the presence of nasals in speech continua in European Portuguese. They were only designed to detect hipernasality and hiponasality productions, respectively.

### **Subjects**

25 adult (12 females) subjects EP native speakers, aged 19 to 27, with no history of otorhinolaryngological disorders, abnormal nasality or without common colds and nasal congestion participated in this preliminary study.

### **Nasometry and the nasometer**

For this study we used the Nasometer 6200-3 da Kay Elemetrics which is a computer-based system, where the oral and nasal acoustic energy components are captured by microphones mounted on each side of a separator plate, that is placed in the subject's upper lip for data collecting.

The nasometer was calibrated before data collecting and the position of the headset was adjusted according to manufacturer's manual specifications in order to get reliable data.

### Experimental procedure

The syllable-repetition test required participants to repeat a consonant-vowel syllable six to ten times in two seconds.

The text reading test required the participants to read a small text with no hesitations or pauses. Each time a participant made a reading mistake or made a longer pause he had to start over and read the text again. For experimental purposes, we only considered the well read texts.

### Results and discussion

Normal nasalance score for stimuli with no nasal segments is in average 10%, both in syllable repetition and in text reading (see Table 1 and Table 2). Nasalance values different from zero in oral stimuli were also reported in studies from other languages (see Hirschberg et al. 2005 for a literature review). Results of syllable types B and C are very near from each other, especially if we consider the high standard deviation values that both present. This may mean that nasal consonants and nasal vowels are quite similar in what acoustic energy is concerned. The highest nasalance score average was achieved in the syllable type D. This value is higher than the ones for other languages with values ranging between 40% and 60%. However, not every other language has nasal vowels in its phonological system so when the values for nasals are shown they are considering only nasal consonants with oral vowels like our syllable-type C.

Nasalance scores for text reading saturated with nasals sounds are considerably lower than the ones registered in syllable repetition test and closer to scores from other languages.

Table 1. Results of nasalance scores by Resonance Syllable type.

Syllable - types	Average	Standard deviation
A – OO	10%	2,83
B – ON	69%	12,22
C – NO	65%	9,74
D - NN	77%	9,45

Table 2. Results of nasalance scores by Text Reading.

Text Reading	Average	Standard deviation
No nasals	10%	3,15
Nasals saturation	44%	8,07

Taking into account the results of this experiment we acknowledge the existence of three levels of normal nasality in European Portuguese: N0 corresponds to a level of nasality of 10% for oral stimuli (syllable and text);

N1 is the intermediate level of nasality - 40% - and corresponding to text reading saturated with nasals; and N2, the highest level, corresponds to 70% of nasalance and characterizes the syllable-type D (nasal consonant and nasal vowel). The difference between each level is of 30% starting from the lowest average nasalance score for oral stimuli.

### Conclusions

This research provided the first global nasalance scores for European Portuguese in a syllable-repetition task and in a text reading task. Results revealed to be quite similar to values from other languages, especially in what oral stimuli are concerned.

Based on these results, we proposed the existence of three levels of normal nasality for European Portuguese: N0, N1 and N2.

Two forthcoming studies, one with 66 EP native children (aged 6 to 10) and the other with 30 females, will provide further evidence for this proposal.

### References

- Anderson, R. T. 1996. Nasometric values for normal Spanish-speaking females: a preliminary report. *Cleft Palate-Craniofacial Journal*, 33, 333-336.
- Haapanen, M.L. 1991. Nasalance scores in normal Finnish speech. *Folia Phoniatrica et Logopaedia*, 43, 197-203.
- Hardin, M. et al. 1992. Correspondence between nasalance scores and listener judgments of hypernasality and hyponasality. *Cleft Palate-Craniofacial Journal*, 29, 346-351.
- Hirschberg et al. 2005. Adaptation of nasometry to Hungarian language and experiences with its clinical application. *International Journal of Pediatric Otorhinolaryngology*, 70, 5, 785-798.
- Mueller, K. et al. 2007. Diagnostic value of nasometry – representative study of patients with cleft palate and normal subjects. *Folia Phoniatrica et Logopaedia*, 59, 219-226.
- Nascimento, F. et al. 1987. *Português Fundamental*. Vol. II, tomo 1, Lisboa: INIC, CLUL.
- Nichols, A. 1999. Nasalance statistics for two Mexican populations. *Cleft Palate-Craniofacial Journal*, 36, 57-63.
- Seaver, E. et al. 1991. A study of nasometric values for normal nasal resonance. *Journal of Speech and Hearing Research* 34, 715-721.
- Van Lierde, K. M. et al. 2000. Nasometric values for normal nasal resonance in the speech of young Flemish adults. *Cleft Palate-Craniofacial Journal*, 38, 112-118.

# **Priming effect on word reading and recall**

Isabel Hub Faria and Paula Luegi

Laboratório de Psicolinguística, LinSe, Centro de Linguística da Universidade de Lisboa

## **Abstract**

This study focuses on priming as a function of exposure to bimodal stimuli of European Portuguese screen centred single words and isolated pictures inserted at the screen's right upper corner, with four kinds of word-picture relation. The eye movements of 18 Portuguese native university students were registered while reading four sets of ten word-picture pairs, and their respective oral recall lists of words or pictures were kept. The results reveal a higher phonological priming effect when recalling words. Results are discussed taking into consideration the eye movements' behaviour values (number and duration of fixations, and number of transitions between word and picture).

Key words: priming, single word reading, word-picture relation, recall

## **Introduction**

The present study is part of a larger research project on reading and comprehension of words, sentences and texts that is taking place at the University of Lisbon. This paper particularly focuses on the possible priming effects of previous reading and recall. In this experiment, priming is a function of simultaneous exposure to bimodal stimuli, since each isolated word is presented, during two seconds, in the centre of a screen, together with a picture inserted on the same screen's right upper corner. This bimodal stimuli exposure is expected to interfere in recall by modality. In other words, our aim is to observe whether the recalling words task is influenced by the images presented simultaneously, and whether the images recall is interfered by the memory of the words read.

For this purpose, we developed an experimental design where the internal relation of each bimodal stimulus, word and image, may be harmonious, consonant or equivalent in terms of reference, or dissonant. In the later case, word and image do not share a complete reference, as category, but only as a subset of the total set of attributes of both, being these attributes of either phonologic or of semantic nature. In view of the foregoing, the priming effect is then considered as the phonological or the semantic nature of the internal relation of each bimodal stimulus.

### Methodology

We establish four kinds of relation for the analysis of the experimental lists of word-picture pairs: (i) synonymy (word: 'mala' (purse); picture: 'mala' (purse)); (ii) word supra and picture infra ordinate (word: 'corpo' (body); picture: 'mão' (hand)); (iii) word infra and picture supra ordinate (word: 'dedo' (toe); picture: 'pé' (foot)); (iv) phonological similarity (word: 'cano' [kanu] (pipe); picture: 'cão' [kãw] (dog)).

Each participant was presented with four sets of ten word-picture pairs and was asked, after each set presentation, to recall the greater possible number of either words read (two sets) or of images visualized (the other two sets).

The eye movements of 18 adult university students Portuguese native speakers were registered with the ASL R6-HS system, while viewing the four sets of ten word-picture pairs. Oral recall lists of words or pictures were kept, together with the respective reading times (word total reading time, image total observation time), number of fixations (word total number of fixations, image total number of fixations) and transitions between word and picture (number of transitions from word to image, number of transitions from image to word).

### Results

A cluster analysis of eye movements' data allows us to observe the existence of three main clusters. The first cluster associates the synonymy word-picture pairs containing words with four or more syllables, and is associated to phonologically related pairs, 50% of the infra-supra and 37,5% of the supra-infra semantic pairs. The second larger cluster associates all the synonymy pairs with three or less number of syllables words and one third of the supra-infra semantic pairs. The third cluster contains all the other left pairs, i.e., 25% of phonological nature, 25% of supra-infra pairs and 37,5% of the infra-supra pairs.

In what concerns synonyms, the results point to a difference in processing based on length of the word. In future experiments, we shall control the word length as well as it's frequency of use. Phonologically related pairs will be also controlled for location of the phonological manipulation (initial, medial or final) or for operation (substitution or adding of a segment).

It should be stressed, though that the first cluster points to higher processing costs and the second cluster to lower ones. The third cluster points to ambiguous situations possibly due to less controlled stimuli construction, namely the pictures' selection.

Analysing the relations among eye movement variables, as shown in Figure1, we observe a modality dependent behaviour: the image variables,

are located in the upper left (1 and 2), and the word measures (6 and 7), and both directions of transitions (4 and 5) are shown at the right hand side.

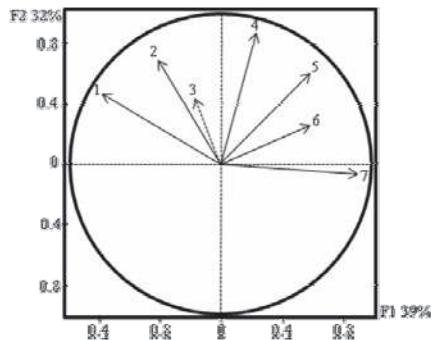


Figure 1. Relations among eye movement variables.

In general, behavioural effects are always registered in the image. The higher is the time spent in the image, the larger is the number of fixations and the number of transitions from word to image. The most retrieved stimulus (3 in Figure 1) are the ones that required higher number of fixations in the image and higher number of transitions from word to image ( $p < 0.05$ ).

Correlations are shown between type of stimulus and eye movement behaviour: phonologically cued pairs have higher values for total number of fixations in the image and higher number of transitions from word to image, when compared to synonyms ( $p < 0.01$ ). They also have a higher total number of transitions from image to word, contrasting with supra-infra semantic pairs ( $p < 0.05$ ).

When recalling words, participants remembered 75% of the words of the phonological pairs. This is statistically significant ( $p < 0.05$ ) when contrasting with the synonyms ones.

When recalling images, there was no significant effect of either phonological or semantic priming.

## Discussion

Thus, our results point out to an effect of phonological priming, not only on eye movements' behaviour but on recall task as well (especially when the task was to remember words). Moreover, our results show that participants make more fixations on the image than on the word, and that this behaviour increased the probability of remembering the stimuli.

In future analyses, these results will be contrasted with works concerning eye movements' when looking to advertisements (see Rayner and Castellano (2007) for a brief review). Since advertisements combine, as typically hybrid



texts (also tested in Faria et al. 2006a, 2006b), images with single words or text, eye movements' behaviour on it may be compared with the kind of stimuli we are using. Accordingly to Rayner and Castellano (2007), when looking to ads, time spent on the image or in the text depends on the viewers' purpose or task; for instance, if they want (or were told that that is the purpose of the study) to buy the product, they spend more time reading the text than viewing the image.

In upcoming experiments, we will, therefore, change the instruction, enlarge the set of stimuli and control for perceptive and semantic properties the images to be used.

## References

- Faria, I. H., Baptista, A., Luegi, P., Taborda, C. 2006a. Interaction and competition between types of representation: An example from eye-tracking registers while processing written words and images. In José Pinto de Lima, Maria Clotilde Almeida e Bernd Sieberg (eds) 2006, *Questions on the Linguistic Sign*, 115-129. Lisboa: Edições Colibri e Centro de Estudos Alemães e Europeus.
- Faria, I. H., Luegi, P., Taborda, C., Baptista, A. 2006b. Recuperação da informação visualizada: interacção e competição entre legendas e imagens. In Oliveira, F. and Barbosa, J. (eds), *Proc. of the XXI Annual Meeting of the Associação Portuguesa de Linguística*, 359-370, Lisboa, Portugal.
- Rayner, K. and Castelano, M.S. 2007. Eye movements during reading, scene perception, visual search, and while looking at print advertisements. In M. Wedel and R. Pieters (eds.) 2007, *Visual Marketing: From attention to action*, 9-42. Lawrence Erlbaum: New Jersey, USA.

# Formulaic expressions in language technology

María Fernández-Parra

Swansea University, United Kingdom

## Abstract

Little attention has been given to the treatment of formulaic expressions in language technology during the past few decades, although such expressions are in fact extremely common both in conversation and in written discourse. Computer-Assisted Translation (CAT) tools are currently the most widely used language technology tools among professional translators. In this paper, I intend to determine the usefulness of such tools in the treatment of formulaic expressions. In particular, the aim is to evaluate the performance of *Trados* (2007 version), the leader in the CAT-tools market, in the treatment of such expressions, compared to the treatment of terms and ordinary translation units, by examining the performance of selected components in the identification and translation of formulaic expressions.

Key words: computer-assisted translation, *Trados*, formulaic language

## Formulaic expressions and *Trados*

Formulaic expressions are understood here as consisting of prefabricated expressions, that is to say “stored and retrieved whole” (Wray 2002: 9) from the mental lexicon at the time of use. Idioms (e.g. *rock the boat*, *spill the beans*), proverbs (e.g. *a stitch in time saves nine*) and collocations (e.g. *auspicious event*, *teething problem*) can be included as subclasses of formulaic language, but formulaic language also includes expressions that would not usually be considered idioms, such as *good morning*, *money talks*, *first thing tomorrow*, etc.

As formulaic expressions are common and pervasive (cf. Jackendoff 1997:156), they constitute challenges for the professional translator, who is bound by constant tight deadlines. Translators may be well versed in the translation of terms, but not to the same degree in the translation of formulaic expressions. An analogy can be established in that both formulaic expressions and terms can be single or multiword units. Since *Trados* has been proven to perform well in the treatment of terms, it could theoretically perform well also in the treatment of formulaic expressions.

## Identification of formulaic expressions

In *Trados*, *Workbench* works as a central platform during translation, drawing on the Translation Memory (TM) and the termbase (TB). Another component of *Trados*, *Multiterm Extract*, has functions that feed into the termbase as well, such as *Monolingual Extraction*. *Workbench* searches the

source text for already existing terms in the *Multiterm* termbase, so that the translator can consult all their previous translations and any associated information about them, whereas the *Monolingual Extraction* tool searches the source text for new terms, so that they may be added to the termbase. Similarly, both identification methods may be used for formulaic expressions.

A selected source text of about 10,000 words was used to determine the usefulness of *Workbench* and *Monolingual Extraction* to identify formulaic expressions. First of all, the text was searched manually and 72 formulaic expressions were identified, with 89 tokens. Then, the search was repeated with both *Workbench* and *Monolingual Extraction* and the results compared.

Only one token per formulaic expression was counted in the *Monolingual Extraction* project. No additional tokens were counted, as formulaic expressions need only be identified once, in order to be included in the termbase. In *Workbench*, however, the identification of each token of a formulaic expression is essential, as the link to its entry in the termbase is only established automatically if identification of the string has taken place. If identification fails, the translator can look up the termbase manually, but at the expense of valuable time.

### **Workbench translation memory**

The main problems encountered by *Workbench* when trying to identify specific segments in a text are variation and noise. In this exercise, two main types of variation were found, inflectional variation and word order inversion. Although fuzzy settings may be altered in *Workbench*, in this exercise the lowest fuzzy settings possible were chosen, in order to obtain the highest number of identified expressions, on one hand, and, on the other hand, in order to try to overcome the problem of variation, which may be more acute for formulaic expressions than for terms.

Overall, the results obtained with *Workbench* are promising. *Workbench* identified 65 tokens of formulaic expressions, out of a maximum of 89. *Workbench* also identified a further 15 tokens of formulaic expressions that had not been identified manually. In total, therefore, 80 tokens were identified with *Workbench*. By contrast, *Workbench* never managed to recognise 13 of the expressions that were found manually.

The most common types of inflectional variation in this exercise were tense variations and singular/plural variations. It would appear that, for *Workbench*, tense variations are bigger obstacles in the recognition of terms/formulaic expressions than singular/plural variations. It recognised all tokens of expressions with singular/plural variations but only half of the tokens displaying tense variations. For example, it recognised the expression *raise money* in the string *First, more money must be raised*, but it did not

recognise the expression *keep pace with* in the string ...*and middle income countries in all regions has not kept pace with the need for expanded....* A possible solution might be to include, as an index field in the termbase, those forms of the expression which differ most from the canonical form, in order to possibly facilitate their recognition by *Workbench*.

As for word order inversion, *Workbench* identified 6 out of the 16 occurrences of expressions with inverted word orders. For example, defying noise, word order and variation, *Workbench* identified the expression make a contribution in the string Funding is only one component of the significant contribution the public sector makes to HIV vaccine....

Noise was another factor to take into account. From the total of 89 occurrences, 30 appeared with varying degrees of noise, out of which *Workbench* identified 16. For example, *Workbench* identified the expression meet a standard from the string In countries that meet public expenditure management standards, aid flows through government... Without taking any of these figures as absolute, they nevertheless indicate that the trend in identifying formulaic expressions with *Workbench* is a successful one overall.

### Monolingual extraction tool

The *Monolingual Extraction* tool differs from *Workbench* in that it produces an editable list of 'candidate' terms, each of which may or may not be selected by the translator for inclusion in the termbase. Because the candidate term strings may contain any portion of source text, not only terms, the *Monolingual Extraction* tool could theoretically be used to identify formulaic expressions 'embedded' in the candidate term.

The extraction settings were set at a minimum of 2 words per expression, since one-word formulaic expressions are not of the same computational interest as expressions of 2 words or more. Further, a maximum of 4 words per expression was also set, having established during the manual search that no expression was longer than 4 words. In a *Monolingual Extraction* project, it is also possible to choose from a wide range of noise level settings, from no noise at all (level 0) to maximum noise (level 1), including every noise level from 0.1 to 0.9 in between. The search for formulaic expressions was performed at each of these levels.

The results obtained with *Monolingual Extraction* were also promising. From a total of 72 formulaic expressions found manually, the program identified 59, and an additional 18 which had not been identified manually. Like *Workbench*, *Monolingual Extraction* did not recognise the same 13 expressions that were only identified manually. With a total of 77 expressions identified, therefore, the *Monolingual Extraction* tool identified more formulaic expressions in total than were identified manually.

The main obstacle in the identification of formulaic expressions appears to be noise rather than variation, as formulaic expressions were identified regardless of inflectional variation and word order inversion. More noise means that more formulaic expressions can be identified, but more time will be needed to search through long lists of candidate terms. Without noise, very few formulaic expressions, and indeed terms, were identified by the machine. In fact, in order to identify a decent number of formulaic expressions, the best noise settings turned out to be 0.8.

### Conclusions

From the discussion in the previous sections, it can be initially concluded that both *Workbench* and *Monolingual Extraction* perform as well in the identification of formulaic expressions as they already do in the identification of terms. Both *Workbench* and *Monolingual Extraction* recognised the majority of formulaic expressions in the source text and both tools identified tokens of formulaic expressions that had not been identified manually.

The best results in the identification of formulaic expressions with *Workbench* are obtained with the lowest possible fuzzy settings, which may not always be the case with terms. Also, variation appears to be a bigger obstacle than noise for *Workbench* to recognise certain formulaic expressions. A possible solution to consider might be the inclusion of additional index fields to certain entries in the termbase containing inflected forms of the expression that differ considerably from the generic form.

The best results in the identification of formulaic expressions with *Monolingual Extraction* were obtained with the higher levels of noise. In these settings, a large amount of formulaic expressions were identified. However, in order to get good results, the translator has to search through long lists of candidate terms. A possible solution to contend with the large amount of noise would be a compromise, namely, to spend the time searching through the long lists of candidate terms only when the number of occurrences of formulaic expressions in a given source text seems to warrant it.

To sum up, *Trados* offers very different functions that can be used in the identification of formulaic expressions, as well as terms. The combined results obtained both with *Workbench* and *Monolingual Extraction* can support the translator significantly.

### References

- Jackendoff, R. 1997. *The Architecture of the language faculty*. Cambridge, MA: Newbury.
- Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: CUP.

# Continuation tunes in two central varieties of Italian: phonetic patterns and phonological issues

Rosa Giordano

Department of Linguistic and Literary Studies, University of Salerno, Italy

## Abstract

This paper presents the phonetic analysis of the intonational patterns conveying the pragmatic contents of *continuation* and/or marking syntactic relations in Rome and Perugia Italian: data show the type and the distribution of accents and boundaries occurring in spontaneous task-oriented dialogues. Results imply considerations about issues concerning systemic, phonotactic and realizational aspects of the intonational system of Italian: the co-presence of accents and boundaries marking continuation contents; their distribution in the intonation group; the phonetic features correlated, in Italian, to phonological contrast and to phonetic variation.

Key words: intonation, Italian, continuation, non-finality, spontaneous speech

## Introduction

The intonation of Italian can vary according to diatopic factors, as it has been shown in several studies following different theoretical and methodological frameworks. Standard Italian and some regional varieties share rising shapes of either accents or boundary tones related to *non-finality*, even if consistent differences can occur among diatopic varieties (Avesani 1996; Gili Fivela 2004; Grice *et al.* 2005; Savino *et al.* 2006). Although the similarity of several melodic patterns in different varieties of Italian has been suggested basing on auditive or phonetic and instrumental analyses, the presence of both continuation accents and boundaries or of structural similarities in their properties is not clearly pointed out yet.

This work provides a preliminary description of continuation tunes for two regional varieties still scarcely investigated: Lazio and Umbria Italian. It is part of a contrastive research on question tunes and continuation tunes and it is based on the analysis of task-oriented dialogues (Giordano 2004, 2006).

Continuation tones are related to two main classes of facts: syntactic relations; pragmatic and conversational factors, which could be also defined *textual*. In the corpus here examined, they can occur in internal position of the dialogic turns, at the edges of intonation groups corresponding to syntactic phrases or clauses which are linked to the following ones by coordination, subordination or juxtaposition. But they can also occur at the edges of dialogic turns, as a means by which the place for turn-taking is signaled and by which the speaker usually can give feedback.

### Method and corpus

Two dialogues, DGtdB04R (Rome) and DGtdA04O (Perugia), were selected from the Italian national corpus CLIPS ([www.clips.unina.it](http://www.clips.unina.it)); they were performed by the same speakers (4) who acted other dialogues analysed in a previous work (Giordano 2006). Dialogic turns were segmented into prosodic groups, basing on phonetic-acoustic criteria (final lengthening, prosodic cohesion and general trends of  $f_0$  and energy) and uditive parsing (Ladd 1996, Hirst and Di Cristo 1998, Kohler 2006). Raw  $f_0$  curve was analysed with Praat ([www.praat.org](http://www.praat.org)) and manually annotated using an INTSINT-like system of transcription;  $f_0$  values and synchronization of the labelled points with segments were then evaluated. Prosodic groups related to continuation functions were selected (118 cases): continuation accents and boundaries are all located at the rightmost edge of the prosodic groups.

### Results and discussion

Continuation accents are similar in both varieties. Their shape consists of a rising movement, which can be phonetically represented as a tonal sequence LH; the *high* (H) target is usually set at the rightmost edge of the nucleus of the last rhythmical strong syllable in the prosodic group; the *low* (L) target is generally placed on one of the preceding syllables. Anyway, other accent types belonging to the declarative series of the intonation inventory of Italian have been found to occur in final position of the prosodic group.

(1) shows an example of the LH accent, associated with the syllable *ET*:

- (1) DGtdB04R\_p1#140:           io ce ne ho una, è una lineETta  
I have one of them, it is a short line

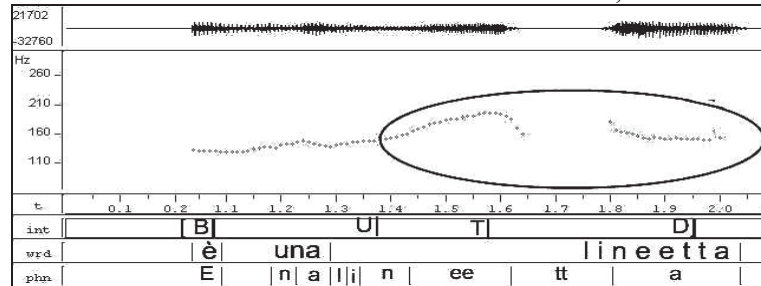


Figure 1: LH accent. Turn: DGtdB04R\_p1#140.

Two different phonetic realizations of the boundary can occur (table 1, figures 2 and 3): 1) a rising movement showing a phonetic tonal sequence LH, or 2) constant values on high levels of the  $f_0$  (a *plateau*), here represented as H. The phonetic context determining the selection of the boundary shape is the preceding pitch accent, in particular the last tonal target of its tonal configuration. When the LH continuation accent is selected, the H boundary occurs, irrespective of the number of syllables



intervening between the accented syllable and the last weak syllable. When other accentual shapes, ending with a L target, are selected, the LH boundary occurs: in fact, continuation boundaries are not necessarily preceded by a continuation accent LH. Both shapes play the same function and this would lead to consider them phonetic variants of the same tone H%.

Figure 2 and 3 are examples of the LH boundary and of the H boundary; in figure 3, the last rhythmical strong position *OC* also carries a LH accent.

- (2) DGtdA04O\_p2#130: c'è un trapezio ... però tronca*TO*  
there is a trapezium ... but truncated
- (3) DGtdA04O\_p1#183: due *OCCHI*  
two eyes

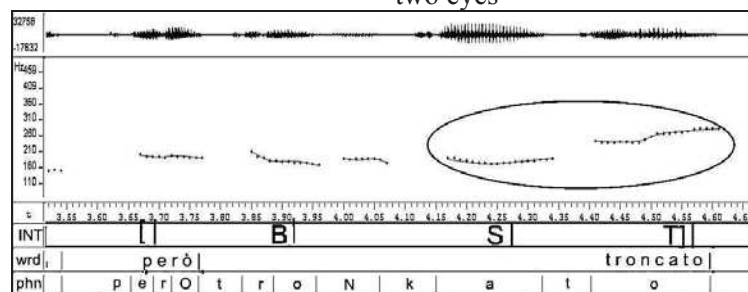


Figure 2: LH boundary. Turn: DGtdA04O\_p2#130.

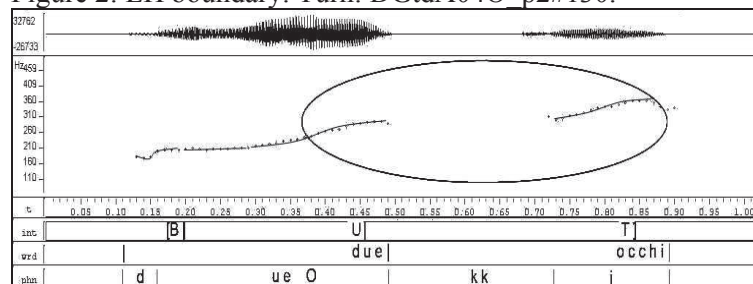


Figure 3: H boundary. Turn: DGtdA04O\_p1#183.

Table 1 and table 2 show the total number of accents and boundaries the mean pitch span in semitones of the LH accents and of the LH boundaries.

Three combinations of the two prosodic devices are found, all of them being used in both varieties. An intonation group can be marked by 1) a continuation accent (11% out of the total number of prosodic groups, 118), or 2) a continuation boundary (53%), or 3) both continuation accent and continuation boundary (36%). The selection of the nuclear pitch accent is then independent of the boundary type: continuation accents and continuation boundaries would not be mutually constrained.

Table 1: Number of accent and boundary types.

Types	LH accent	LH boundary	H boundary	Total
Number	56	62	43	161



Table 2: LH accent and LH boundary: mean excursion size in ST.

Span in ST	Rome (tdB04R)		Perugia (tdA04O)	
	Speaker 1	Speaker 2	Speaker 1	Speaker 2
LH accent	4.8	3.6	6	5.3
LH boundary	5	4.4	7	5.8

## Conclusion

Rome and Perugia Italian show consistent similarity among the phonetic tonal sequences that continuation accents (LH) and boundaries (LH or H) present; also contextually constrained variations of their tonal forms systematically occur. Despite of such evidences, the phonological representation of these units for these varieties is not unproblematic, because question tunes show phonetic shapes and properties very similar to the ones found for continuation tunes (see Giordano 2004, 2006): LHL accent (y/n questions) and LH accents (asking about new information); LH and H boundaries. Such topics involve considerations about more general aspects of the theoretical explanation of melodic phenomena: the role of phonetic details in determining phonological contrast; the entity and the kind of their variation - gradient or categorical - among diatopic varieties of a language.

## References

- Avesani, C., 1996. TOBIT: un sistema di trascrizione per l'intonazione italiana. In Lazzari G. (ed.), *Atti delle XI Giornate di Studio del GFS*, 85-98.
- Gili Fivela B. 2004. The phonetics and phonology of intonation: The case of Pisa Italian. PhD Dissertation. Scuola Normale Superiore, Pisa.
- Giordano, R. 2004. *Aspetti strutturali e interrelazioni contestuali dell'intonazione dell'italiano*. Phd dissertation. University of Perugia, Italy.
- Giordano, R. 2006. The intonation of polar questions in two central varieties of Italian. In Hoffmann R. and Mixdorff H. (eds.), *Proc. of Speech Prosody 2006*, Dresden, Germany. DVD. Dresden, TUD press.
- Grice, M., D'Imperio, M., Savino, M., Avesani, C., 2005. Towards a strategy for labelling varieties of Italian. In Jun S.A. (ed.), *Prosodic Models and Transcription*, 55-83. Oxford: OUP.
- Hirst, D., Di Cristo, A. (eds.) 1998. *Intonation Systems*. Cambridge: CUP.
- Kohler, K.J. 2006. Paradigms in Experimental Prosodic Analysis. In Sudhoff S. et al. (eds.), *Methods in Empirical Prosody Research*, 123-151. Berlin and New York: Walter de Gruyter.
- Ladd, D.R. 1996. *Intonational phonology*. CUP.
- Savino, E., Grice, M., Gili Fivela, B., Marotta, G. 2006. Intonational cues to discourse structure. In Hoffmann R. and Mixdorff H. (eds.), *Proc. of Speech Prosody 2006*. Dresden, TUD press.

# **Model-based duration analysis on English natives and Thai learners**

Chatchawarn Hansakunbuntheung<sup>1</sup>, Hiroaki Kato<sup>2</sup> and Yoshinori Sagisaka<sup>1</sup>

<sup>1</sup>GITI/Language and Speech Science Research Labs, Waseda University, Japan

<sup>2</sup>NICT/ATR Cognitive Information Science Labs, Japan

## **Abstract**

This paper presents a model-based statistical duration analysis on the comparison between English natives and Thai-native English learners. The analyses were carried out to characterize non-native's duration control on (1) control differences between natives and learners and on (2) the relationship between the duration characteristics and learners' background reflecting English experiences. The analyses showed lower speech rate and strong influence of the first language characteristics in beginner's speech and gradual improvements according to their English experiences quantitatively.

Keywords: English duration, computational duration modeling, second language

## **Introduction**

Timing control is one of the most fundamental and essential issues in spoken language education and in smooth speech communication. For English learners, differences in timing control characteristic should be clearly understood and self-evaluated to improve their own English-speaking skill. Thus, we need a scientific quantitative evaluation method of English timing and the knowledge of timing differences based on the model for the improvement of non-native English learner's timing control skill.

In this paper, we proposed a method of model-based statistical analysis on timing characteristics for Thai-native English learners. The analysis was carried out by comparing durational differences between learners and English natives. In the analysis, we quantified the differences in speech rate and segmental duration using a statistical model. Furthermore, the prediction errors between another statistical model and Thai learners were employed to associate with learners' English experiences quantitatively.

## **Model-based duration analysis**

For learner's proficiency evaluation in timing control, we adopted an objective measure of an average difference between each segmental duration of a target learner and native's one instead of conventional subjective measures. To be free from subject dependent durational characteristics, a statistical duration model was trained using English speech data uttered by multiple native speakers. We measured the durational differences between

statistically predicted values, and, used the resultant prediction error as a measure to evaluate proficiency in English timing control.

### **Specifications on database**

**English speech database.** We used three English speech databases. The first one was the ARCTIC database (Kominek 2003) read by English-speaking natives. The database was separated into two sets: set A of 593 sentences, and, set B of 539 sentences. The second database was a read English speech database of the fairy tale “The north wind and the sun” from CUCHLOE corpus that read by English natives, and, speakers from English-as-an-official countries. The third one from NECTEC was a speech database of the fairy tale that read by 45 Thai learners of English, and, one Indian English speaker.

The above databases were grouped into four sets for modeling and analysis. The first set consisted of speech data from ARCTIC set A uttered by four US speakers. This set was used as a training set for the reference English duration model. The second set consisted of speech data from ARCTIC set B by the same four speakers. We referred this set as a (speaker) close set, and, used to evaluate consistency of the model. To evaluate the validity with various English accents, we made an open set. The third set contained 3 non-US-accent speakers from CMU ARCTIC set B, 7 speakers from CUCHLOE, and, one Indian English speaker from NECTEC. The last set contains 45 Thai learners of English from NECTEC. This set was used as a test set to evaluate English duration characteristics of the Thai learners.

**Learner’s information.** For Thai learners, the learner’s information related to English experience were collected. It consists of age (years), English-educated period (years), educating period in English-as-an-official-language countries (years), and, average duration for English usage in a week (hours).

**Factors for statistical modeling.** For a statistical model, the following categories of control factors were employed; current and four context phones, stress, phone position and numbers of constituent phones in syllable, word and phrase, syllable position and numbers of syllables in word and phrase, narrow and board parts of speech.

**A statistical model for English duration.** Before modeling, we normalized phone duration for each speaker using z-score with mean and standard deviation to eliminate speech rate effect for inter-speaker comparison. For the modeling, we adopted a linear regression based on categorical factors (Hayashi 1950). In this model, each sub-category of the factor categories was encoded as “1” if the considering sub-category exists in current segment. Otherwise, it was encoded as “0”. By adopting least-square-error minimization criteria, the model coefficient representing the contributions of the control factors were calculated. Finally, root-mean-squared errors of

differences between measured and estimated duration were calculated for each speaker for comparison.

## Experimental results

### Prediction error analysis on segmental duration

An average of duration prediction errors was employed as an objective measure for the proficiency in English timing control. Figure 1 shows noticeable grouping of prediction errors by speaker profiles. For the close set consisting of the same speakers employed in model training, the errors showed the closest distance from the error of the training set. The close set also showed the lowest errors in all groups of speakers. Thus, the results showed consistent accurate prediction of the model on both training and close sets. For other native speakers, their prediction errors were much close to ones of speakers used in model training and smaller than most of Thai learners. The learners living in English-as-an-official-countries for more than 10 years showed salient decreasing of the distances from the reference model. Furthermore, the learners having no experience in English-as-an-official-language country showed bigger prediction errors with large variation of English skills.

### English duration characteristics of the Thai learners

In Figure 2, “S”, “I”, “AI”, “M”, “BF”, “F” labels represent syllable position in monosyllabic word, initial, after-initial, mid, before-final, and, final position in a word, respectively. Big errors of the Thai learners were found at the ends of a word and a phrase. This is resulted from the difference in stress placement between Thai and English. In Thai, primary stress syllable is always located at the last syllable of words, while, stress placement in English need not to be the case.

### Analysis based on speaker’s English learning experience

By observing from the model coefficient of the model between the prediction errors and the experience, Table 1. shows a noticeably corresponding trend between Educating period in English-as-an-official-language country and the errors. It shows gradual improvements according to their English experiences when the period more than five years. This result corresponds with the results in Figure 1.

Table 1. Model coefficients of the Educating period control factor.

Period (years)	0	0<Y≤ 5	6≤Y≤ 10	11≤Y≤ 15
Model coefficient	0.506	0.570	0.0	-0.780

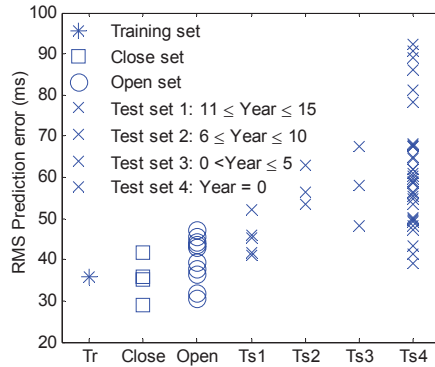


Figure 1. Comparison of prediction errors between English natives and Thai learners with different educating period in English-as-an-official-language countries.

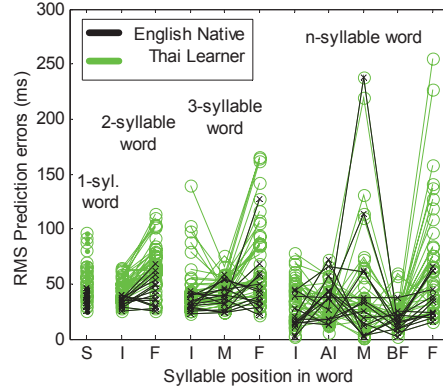


Figure 2. Prediction errors comparing timing control of syllable position in word with different word lengths between the native and the learners.

## Conclusions

We proposed a method to analyze English duration characteristics of Thai learners of English for the proficiency evaluation of English timing control. Experimental analyses on duration prediction errors showed the effectiveness of the proposed objective evaluation using statistical duration characteristics based on the English duration model as a reference. The results suggested the use of global prediction errors, prediction errors from timing controls, and, the learners' background as objective measures. These results could lead to further study on quantitative evaluation of English skill.

## Acknowledgements

We would like to thank our colleagues at National Electronics and Computer Technology Center (NECTEC), for collecting the English speech data of Thai learners and native. We also would like to express our special thanks to Prof. Helen Meng (Chinese University of Hong Kong) for providing the CUCHLOE English speech corpus. This work was supported in part by Waseda University RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for Scientific Research B, No. 20300069 of JSPS.

## References

- Kominek, J, and, Black, A. W., 2003, CMU ARCTIC database for speech synthesis.
- Hayashi, C. 1950. On the quantification of qualitative data from the mathematic-statistical point of view. *Annals of Institute of Statistical Mathematics*, vol. 2.

# Lexicalization of natural actions and cross-linguistic stability

Paul E. Hemeren<sup>1</sup>, Sofia Kasviki<sup>1</sup> and Barbara Gawronska<sup>2</sup>

<sup>1</sup>School of Humanities and Informatics, University of Skövde, Sweden

<sup>2</sup>Department of Foreign Languages and Translation, University of Agder, Norway

## Abstract

To what extent do Modern Greek, Polish, Swedish and American English similarly lexicalize action concepts, and how similar are the semantic associations between verbs denoting natural actions? Previous results indicate cross-linguistic stability between American English, Swedish, and Polish in verbs denoting basic human body movement, mouth movements, and sound production. The research reported here extends the cross-linguistic comparison to include Greek, which, unlike Polish, American English and Swedish, is a path-language. We used action imagery criteria to obtain lists of verbs from native Greek speakers. The data were analyzed by using multidimensional scaling, and the results were compared to those previously obtained.

Key words: motion verbs, natural actions, cross-linguistic stability, manner, path

## Introduction

Human actions represent a conceptual domain where motion plays an important role in category structure and lexicalization as well as recognition. Learning to perceptually discriminate among various actions requires access to the spatiotemporal patterns in human movement (Giese and Poggio 2003). It seems that the reverse is also true: retrieving natural actions from memory requires access to spatiotemporal features associated with specific actions.

If perceptual criteria play a fundamental role in the lexicalization of action concepts, then we should see some cross-linguistic stability in mental lexical representations despite the fact that previous research (Vinson and Vigliocco 2002) has shown that the semantic space for verbs contains less well-defined boundaries in contrast to the semantic space for nouns.

In previous research, Hemeren (1996) and Hemeren and Gawronska (2007) asked American, Polish, and Swedish informants to spontaneously list verbs denoting actions that can be easily recognized when seen and can be visualized as a mental image. The results showed a clear tendency for informants to list verbs for basic level actions first. Verbs for subordinate level actions were mentioned later and less frequently. In all three languages investigated, we found significant correlations between total verb frequency (TF) and the mean ordinal position (MOP). By using multidimensional scaling as a technique to investigate the structure of semantic associations

within a language, further analyses revealed a similar tendency for the three languages: basic motion verbs ‘run’, ‘walk’, ‘jump’ and verbs referring to mouth motion and/or sound production: ‘laugh’, ‘talk’, ‘sing’ had a similar organization in terms of derived distances in the three-dimensional semantic space. These results indicate a tendency towards cross-linguistic stability between motion verbs and to a certain extent the mouth motion/vocal verbs.

### The current study

The research presented here compares data coming from speakers of “manner languages”, like English, Swedish, and Polish, where the manner of motion is normally encoded in the semantics of the main verb, and from speakers of “path languages”, like Greek.

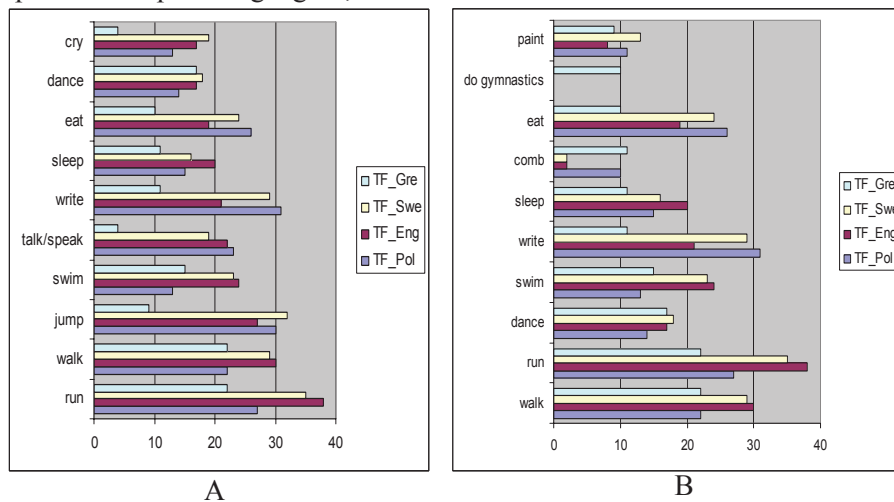


Figure 1. Panel A shows the 10 most frequent English verbs compared with the frequencies of the corresponding verbs in Greek, Polish, and Swedish. Panel B shows the 10 most frequent Greek verbs compared with the frequencies of the corresponding verbs in English, Polish, and Swedish.

We applied the same method as in the previous investigations. Thirty-five native speakers of Greek were asked to spontaneously list verbs denoting actions that are easy to imagine. The data were first analyzed with respect to verb frequency, list position, and the correlation between those two measures. Some results concerning frequencies are shown in Figure 1.

One major difference between the Greek verb lists and the lists from the other languages was the fact that the Greek lists contained significantly fewer words per list than the other lists, all  $ps < .001$ . The mean number of words per list for English was 37, Swedish 41, Polish 33 and Greek 18.



For the Greek informants, as for the other language groups, the more frequently a verb occurred on the lists, the earlier it also appeared, Pearson's  $r = -.48, p < .05$ . This suggests that highly frequent verbs are easily accessed and therefore are produced early on during the task.

There was also a clear tendency to list basic motion verbs and verbs referring to basic human behaviours ("eat", "talk" etc. – see Table 1). The correlation between the four languages regarding listed verbs was significant for all pairwise language comparisons, all  $p$ s  $< .005$ , except for Polish and Greek. In this last case, the Pearson correlation coefficient was  $.26, p > .20$ .

Table 1. 10 most frequently listed actions in the four languages. Verbs in bold face were found among the top ten entries in all four languages; verbs found in three of the four languages are marked by italics.

Am. English	Greek	Polish	Swedish
run	run/ τρέχω	write/ pisać	run /springa
walk	walk/ περπατάω	jump/ skakać	jump/hoppa
jump	dance/ χορός	read/ czytać	write /skriva
swim	swim/ κολύμπι	run/ biegać, bieć	walk /gå
talk/speak	write/ γράφω	eat/ jeść	eat/äta
write	sleep/ ύπνος	drink/ pić	swim/simma
sleep	comb/ χτένισμα	talk/ mówić	laugh/skratta
eat	eat/ μασουλάω	walk/ chodzić, iść	cry/gråta
dance	do gymnastics/ γυμναστική	sing/ śpiewać	talk/speak
cry	jump/ πηδώ paint/ μπογιατίζω	wash (oneself)/ myć (się)	kiss/kyssas

Twenty-five of the most frequent verbs from the Greek lists were used to create a semantic space by using multidimensional scaling. Figure 2 shows the 3-dimensional solution for these verbs. As can be seen there is some tendency for vocal and mouth related actions to be close to one another. Verbs of bodily movement are also close to one another.

Correlation analyses were performed on the Euclidean distances between all 300 verb pairs for all 4 languages. With the exception of one correlation, the other 5 correlations were, although significant, ranged from  $r = .18$  to  $.37$ . A larger correlation was obtained for the comparison between Swedish and Polish,  $r = .56, p < .001$ . This indicates that the network of semantic associations is similar for Swedish and Polish given the verbs included in the analysis. Noteworthy however is the finding that the semantic space for the Greek data did correlate with any of the other three languages.



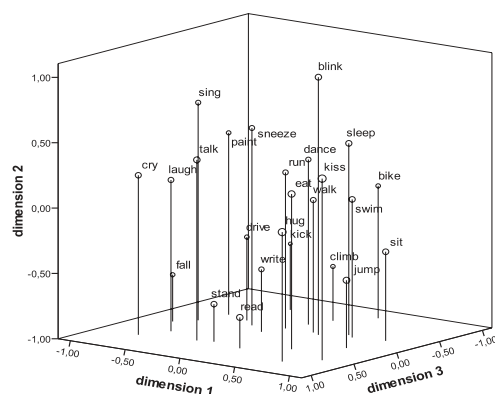


Figure 2. Euclidean distances between the Greek action verbs.

## Conclusions

The limited number of words listed by the Greek informants is a factor that reduces the possibility of drawing reliable conclusions regarding the difference in lexical verb organisation between path and manner languages. One clear difference between the Greek data and the data from the three manner languages is that the correlation between semantic associations between Greek and the three manner languages appears to be relatively weak. This means that although similar actions are frequently listed by informants in the 4 languages, the associations *between* the verbs seem to be different. Nevertheless, the high frequency of motion and sound production verbs indicates that the cognitive status of natural human motion patterns and human vocal and emotional behaviour corresponds to similar lexical accessibility in all four languages.

## References

- Giese, M.A. and Poggio, T. 2003. Neural mechanisms for the recognition of biological movements. *Nature Review Neuroscience* 4, 179-192.
- Hemeren, P.E. 1996. Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica* 91, 39-66.
- Hemeren, P.E. and Gawronska, B. 2007 Lexicalization of natural actions and cross-linguistic stability. In Ahlsén, E. et al. (eds) 2007, *Communication – Action – Meaning. A Festschrift to Jens Allwood*, 57-74. Göteborg, Sweden, Göteborg University.
- Vinson, D.P. and Vigliocco, G. 2002. A semantic analysis of grammatical class impairments: semantic representations of object nouns, action nouns, and action verbs. *Journal of Neurolinguistics* 15, 317-351.

# **“Deep and raspy” or “High and squeaky”: a cross-linguistic study of voice perception and voice labeling**

Begoña Payá Herrero

Ludwig Maximilians Universität, Germany

## **Abstract**

Vocal correlates for identity markers were sought by interviewing 60 candidates of three nationalities and by analysing their average pitch, intensity, speech rate and speech length. National differences were found in pitch and specially speech rate but no clear vocal correlates for the personality parameter of introversion/extroversion were found. An analysis of the voice labels given by the candidates showed that there are common voice stereotypes for sexy voices and unpleasant voices, despite some minor national differences in the use of labels such as *sweet* or *nasal*. Finally, a perception test proved that there is a high-interrater consistency in judgements about voice and personality and that basic emotions and attitudes are well recognised, irrespectively of the language knowledge.

Key words: vocal correlates, identity markers, voice perception, voice stereotypes, voice labels.

## **Introduction**

It is through the combination of *voice dynamics* aspects and *voice quality* aspects (Abercrombie 1967:7) that indexical information about the speaker's identity is conveyed through voice. Indexical information can be said to be either *biological*, *psychological* or *social* (Laver, 1991: 154) or from the perspective of the vocal cues that give this information, we can speak about *physical markers*, *psychological markers* and *social markers* (Laver & Trudgill, 1979:2). When it comes to judgments people make about the three kinds of indexical information mentioned when hearing a voice, *psychological* and *social information* judgments tend to be less accurate, just because these are the ones that are culturally based, arbitrary and learnt. Yet these are the ones upon which I will be focusing my research, since they are the “fingerprints” of our emotions, attitudes and personality. A starting point for my project was to collect voice labels that non-experts use that would include young female speakers of Spanish, German and English.

## **Interviews**

My target group was made up of female speakers between 20 and 30 years of age: 20 Germans, 20 Spanish, and 20 Americans. Face-to-face individual interviews were carried out in each of their mother tongues and they had to answer a total of 11 questions.

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.

### Voice labels

Using Laver's distinction (1974), voice labels provided by candidates in their answers were mainly classified into **phonetic labels** and **impressionistic labels**. The following interesting observations were made:

- In describing their own voice and sexy voices, there was a clear tendency for candidates of the three nationalities to choose *phonetic labels* more often than *impressionistic labels*, with *voice dynamics* and pitch labels, being the most frequent ones.
- In the description of unpleasant voices there was again a clear preference amongst the three nationalities for *phonetic labels*, but this time *voice dynamics* labels and *voice quality* labels were equally preferred. Two adjectives were found to be the most frequent in all three nationalities: *high* and *squeaky*, which were the stereotypes for unpleasant voices.
- A stereotypical sexy voice seems to be the *deep and raspy* kind in the case of all the languages studied. This then would be, at least for western culture, to which the languages studied belong, the perceived factors pointing to sexiness in voice. This same idea can be found in Rodero's (2001: 8-9) study of pleasant voices in radio news.
- Another interesting observation is the perception of *nasality* as unpleasant only for the German and American candidates. *Nasal* was only used once by a Spanish speaker, whereas, it was used by three different Germans and by five different Americans. This seems to show that only in the phonological systems where nasality is strongly present (at least here for American English and German), can this aspect be perceived as unpleasant.

### Voice, nationality and personality

Candidates interviewed also answered a personality test, the so-called Jung Myers Briggs Typology test in order to classify them according to the introversion/extroversion parameter of personality. All the voice samples from the interviews were then analyzed phonetically with Praat (length, intensity, frequency) and the average speech rate in SPM was calculated. The phonetic results were then contrasted in terms of nationalities and personality groups (introverted vs. extroverted) to see if there were on the one hand big national differences and on the other hand to search for clear correlations between specific vocal cues and the degree of introversion/extroversion. Certain national differences were found in speech rate (spanish women being the fastest speakers) and in pitch (spanish women had the lowest average pitch, german women the highest), but not in intensity. See figures 1 and 2.

However, no obvious correlation could be found between the degree of introversion/extroversion and particular trends in the use of voice cues (such as expected high speech rate for extroverts or lower intensity for introverts). Extroverts did speak longer in average throughout all nationalities but not all personality groups displayed general tendencies in pitch, intensity or speech rate. This seems to indicate that introversion or extroversion is not clearly manifested by a single voice parameter.

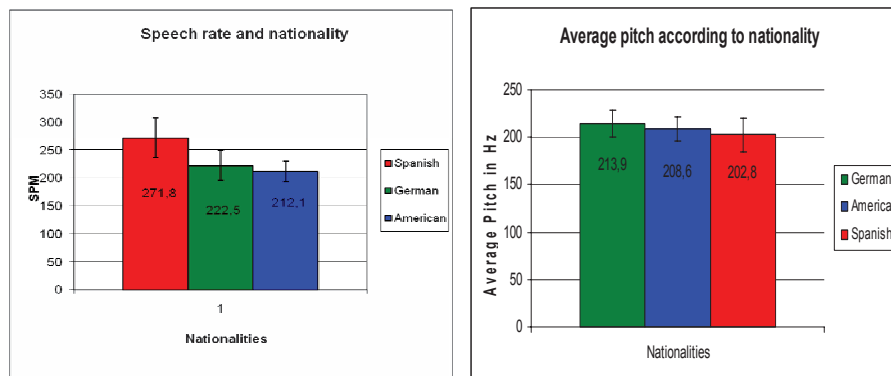


Figure 1. Speech rate and nationality. Figure 2. Pitch and nationality

### Perception test

The next step was to move on from the phonetic analysis and the label classification to the perception of voice by non experts. The perception test was devised by using differentiated female voices from the interviews and acted emotions and attitudes as stimuli and the labels from the classification in a bipolar five point scale, inspired by Osgood's (1964) semantic differential technique. A total of four different groups of voice judges took the test: some understanding the target language other not. All the answers given by a total of 40 voice judges were then classified in an Excel table that gave the average results for each stimulus heard.

In general, in terms of judgments on voice dynamics and voice quality, the results showed that tone and pitch are the easiest recognized. Regarding the judgments on personality, there was a correlation between voices felt as self-confident/strong/adult and extrovert. Voices perceived as emotional were also felt to be aggressive and squeaky. Similarly there was a correlation between fast and active voices. In terms of the judgments on synesthetic labels, voices perceived as high, were all judged to be bright and light, and voices seen as warm were also in general soft. This all proves that there is high interrater consistency in the answers.

Moreover, the initial stereotypes found in the interviews claiming that in general high voices are unpleasant, whereas deeper voices tend to be attractive, was also confirmed in the perception test.

When it came to the perception of emotions and attitudes, most candidates, irrespectively of their knowledge of the language heard, made right guesses.

This then supports the idea that attitudes and emotions are transmitted more by different vocal cues than by the content level of language, at least for the languages studied.

### **Conclusions**

Within a relatively small scope of cross-linguistic samples, the present study has perhaps firstly shed some light upon how voice perception is mirrored by voice labels. There are voice stereotypes that are not purely language dependant, but rather, common to the Western culture they belong to. However, a small number of language-dependant voice labels have also been found in this study. Secondly, the perception test carried out echoes many other projects that have stated how voice is a key transmitter of indexical information, by showing that there is something we could call “indexical stereotypy of voice”: in other words, that people often do make interpretations of indexical information based on vocal cues, whether what is inferred is accurate or not. Thirdly, the crosslinguistic perception test has helped support the idea that it is the phonetic level of language -and not the content level- that conveys most of the indexical information, and that it helps people make accurate guesses about emotions and attitudes.

### **Acknowledgements**

I thank my two PhD tutors, Dr. Pennock and Prof. Schulze for their supervision and I thank the Mutua Madrileña for their economic support.

### **References**

- Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago: Aldine.
- Laver, J. 1974. Labels for voices. *Journal of the International Phonetic Association* 4: 62-75
- Laver, J. and Trudgill, P. 1979. Phonetic and linguistic markers in speech. In Scherer, K.R. and H.Giles, H. (eds.), *Social markers in speech*, 1-32. Cambridge: Cambridge University Press.
- Laver, J. 1991. *The Gift of Speech. Papers in the Analysis of Speech and Voice*. Edinburgh: Edinburgh University Press.
- Rodero, E. 2001. El tono de la voz masculina y femenina en los informativos radiofónicos: un análisis comparativo. Valladolid: Congreso Internacional de Mujeres, Hombres y Medios de Comunicación. Retrieved 11/03/2007 from [www.bocc.ubi.pt/pag/rodero-emma-ono-voz-femenina.pdf](http://www.bocc.ubi.pt/pag/rodero-emma-ono-voz-femenina.pdf)
- Osgood, C. 1964. Semantic Differential Technique in the Comparative Study of Cultures. *American Anthropologist* 66/3, 171-200.

# The effect of focus on lexical tones in Vietnamese

Stefanie Jannedy

Center of General Linguistics (ZAS), Berlin, Germany

## Abstract

We investigated the effect of focus on the formation of  $f_0$  in two of the six tones of the Hà Nội dialect of Vietnamese. This dialect contrasts six lexical tones, has no tone-sandhi, the language does not have focus markers and uses prosody exclusively to express pragmatic contrasts (Jannedy, 2007) by means of intonational emphasis in ways similar to English or German. To tease out the effect of focus on the duration and  $f_0$  of the level tone *ngang* and the rising tone *sắc*, identical short SVO-type utterances were elicited with a question-answer paradigm under three focus conditions (subject-, verb-, and object focus). Results indicate that the tone especially under subject- and verb- but also object focus shows an  $f_0$ -excursion for the *ngang* tone and a scoop and sharp rise for *sắc*. Duration of the focused syllable also increases.

Key words: lexical tone, Vietnamese, focus, intonation

## Introduction

Mon-Khmer languages are known for the complexity of their tone system: lexical contrasts are marked by tonal (pitch) as well as laryngeal features (Yip, 1995). This interaction of voice quality and lexical tone also characterizes Vietnamese (Brunelle, 2003). Several studies have explored the six-tone system of the northern (Hanoi) Vietnamese dialect and have established that there is a higher and a lower pitch register (Nguyễn and Edmondson, 1997). A graph of the shapes of the six lexical tones representative of the standard Hà Nội dialect can be found in Jannedy and Brunelle (2007). Three of the six tones are produced in a modal voice. In this work we will primarily describe the behavior of the rising tone *sắc* which canonically rises from the bottom of the pitch range to the top, and the neutral tone *ngang* which remains fairly stable in pitch throughout in its canonical shape. (In the larger study, we also investigated the low-level tone *huyền* and all positional combinations of the three modal tones to get an understanding of their behavior under focus.)

Brunelle (2003) investigated the effect of coarticulation on neighboring tones in Hà Nội Vietnamese and found that progressive assimilation is stronger than anticipatory assimilation. Michaud and Vu (2004) propose based on their results that “[...] a stable correlate of emphasis is curve amplification, manifested [...] as an increased slope of  $F_0$  curve [...] or as  $F_0$  register raising.” Jannedy (2007) tested whether native listeners of Hà Nội

Vietnamese were able to recover the information structure encoded in utterances recorded under different focus conditions and found that listeners were reliably able to identify subject-, verb- and object-focus utterances.

## Method

### Corpus

We constructed a corpus consisting of 21 simple SVO-sentences, all differing with regard to the lexical tones on the subject, verb and object. These sentences were recorded in a question-answer context under laboratory conditions. For an example, see Table 1.

Table 1. Má lấy muối – *Mother takes salt* . – *sắc sắc sắc*.

Chuyện gì thế?	What is happening?	[Má lấy muối] <sub>F</sub>	Sentential
Ai lấy muối đây?	Who takes salt?	[Má] <sub>F</sub> lấy muối	Subject
Má lấy gì đây?	What is Ngà taking?	Má lấy [muối] <sub>F</sub>	Object
Má làm gì với muối đây?	What is Mother doing with the salt?	Má [lấy] <sub>F</sub> muối	Verb
Má làm gì đây?	What is Mother doing?	Má [lấy muối] <sub>F</sub>	VP

Only two of these tonal configurations are described in this paper. Four native Hà Nội Vietnamese speakers (3 females, 1 male) read the sentences 5 times as a reply to questions.

### Measurements

All elicited utterances were constructed to contain three monosyllabic words with either the level tone *ngang* or the rising tone *sắc* on each word. The data was phonemically annotated in Praat (Boersma and Weenink, 2007) and the duration and F0-maximum was extracted for each phoneme. We also calculated the F0 at different time points (begin, 25%, 50%, and 75% into the vowel) within the tone bearing vowel in the word under focus. Also, time-normalized F0 contours were created via a script (Xu, 2007) and overlaid for easier visual comparison.

### Results and discussion

The lines in the graphs below show time-normalized f0-trajectories for subject focus (black), verb focus (dark grey) and object focus (light grey) for the level tone *ngang* (left panels) and the rising tone *sắc* (right panels) for four speakers (F1-F3 and M1). Note that the language has complex vowel sounds much like diphthongs and that each vowel was preceded by a nasal or lateral consonant for easier tracking of the f0.

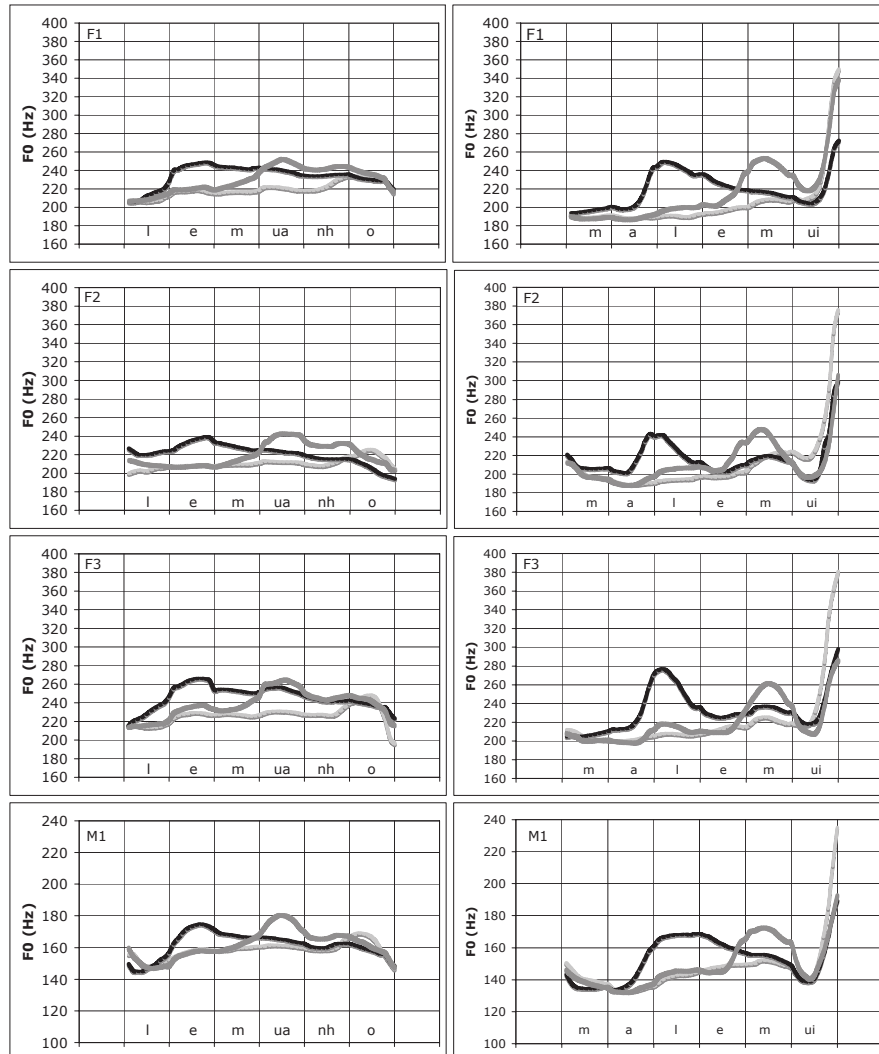


Figure 1. Time normalized F0-contours (Xu, 2007) for the level tone *ngang* (left) and the rising tone *sắc* (right) for three female (F1-F3) and one male speaker (M1).

We fitted linear mixed effects models with the measurements as the dependent variables, the *Focussed Word* as the fixed effect and *Focus Condition* and *Speaker* as random effects, separately for both tonal conditions. All main effects came out as significant ( $p < .001$ ). Table 2 shows the results of the post-hoc comparisons: for example, for the *ngang* tone, in the subject (verb and object) focus condition(s), we compared the values for the focused subject with the subjects in the V and O focus conditions (paradigmatic comparison).



Table 2. Results of post-hoc Tukey-test of Subject-, Verb-, and Object focus conditions with S-, V-, and O in non-focus conditions in the same position.

Tone	<i>ngang</i> (level)						<i>sắc</i> (rising)					
Foc	Subject		Verb		Object		Subject		Verb		Object	
no Foc	V	O	S	O	S	V	V	O	S	O	S	V
F0-max	***	***	**	***	n.s.	n.s.	***	***	n.s.	***	***	**
Dur	***	***	***	***	***	***	***	***	***	***	n.s.	n.s.
F0-start	***	***	n.s.	***	n.s.	n.s.	*	n.s.	***	n.s.	n.s.	*
F0-25%	***	***	n.s.	***	n.s.	n.s.	*	*	***	n.s.	n.s.	n.s.
F0-50%	***	***	**	***	n.s.	n.s.	*	*	n.s.	n.s.	**	**
F0-75%	***	***	**	***	n.s.	n.s.	*	**	n.s.	n.s.	n.s.	n.s.
*** = p<.001    ** = p<.01    * = p<.05    n.s. = not significant												

Results indicate that focus is expressed through *f0* and *duration*: duration increases in most cases significantly under focus. An Object in the O-Focus condition is less often found to be significantly different from objects in on-focus conditions. Duration and *f0* cues may be confounded by final lengthening and lowering effects. For *ngang*, the maximum *f0*-excursion is aligned with the tone bearing unit, while the rising tone shows much of its scooped shape under focus and utterance finally, but not medially unfocussed.

## References

- Boersma, P. and Weenink, D. 2007. Praat: doing phonetics by computer (Version 4.6.36) [Computer program]. (<http://www.praat.org/>).
- Brunelle, M. 2003. Coarticulation Effects in Northern Vietnamese Tones. Proceedings of the 15th ICPhS. (<http://aix1.uottawa.ca/~mbrunell/Research.htm>).
- Brunelle, M. and Jannedy, S. 2007. Social Factors in the Perception of Vietnamese Tones. Proc. of the 16th Intern. Congress of Phonetic Sciences, pp. 1461-64.
- Jannedy, S. 2007. Prosodic Focus in Vietnamese. In Ishihara, S; Jannedy, S. and Schwarz, A. (eds.) Interdisciplinary Studies on Information Structure (ISIS), pp. 209-230, Potsdam University Press.
- Michaud, A. and Vu, T.N. 2004. Glottalized and Non-Glottalized Tones under Emphasis: Open Quotient Curves remain stable, F0 curve is modified. Proceedings of the Speech Prosody 2004, pp. 745-748, Nara, Japan.
- Nguyễn, V. L. and Edmondson, J. 1997. Tones and Voice Quality in Modern Northern Vietnamese: Instrumental Case Studies. Mon-Khmer Studies 28, 1-18.
- Xu, Yi. 2007. <http://www.phon.ucl.ac.uk/home/yi/downloads.html>.

# **MORPHEMIA: a semi-supervised algorithm for the segmentation of Modern Greek words into morphemes**

Constandinos Kalimeris and Stelios Bakamidis

Voice and Sound Technology Department, Institute for Language and Speech Processing (ILSP), Greece

## **Abstract**

The present paper reports on MORPHEMIA, a semi-supervised machine-learning algorithm designed to segment Modern Greek (MG) words into morphemes. The algorithm segments its input iteratively. During its first iteration, the algorithm uses its a priori linguistic knowledge. At the end of each successful iteration, the algorithm extracts new morphological knowledge which is utilised during its next iteration. Thus, with each successful iteration, the algorithm segments an increasing amount of its input data. The algorithm uses a metric to decide whether a given extracted piece of morphological knowledge will improve its performance and only accepts it if it will. Thus, its output gradually improves in quality. MORPHEMIA terminates its operation when new knowledge can no longer be extracted from its input data.

Key words: machine-learning, morphologically rich languages, statistical language modeling, morphological lexicons

## **Introduction**

Recently, the idea of building statistical language models based on linguistic units smaller than the word has been gaining ground (Kirchhoff and Sarikaya 2007). The advantages of utilising a unit such as the morpheme in language technology applications handling linguistic data of Morphologically Rich Languages (MLRs) may seem obvious. However, morphemes are notoriously difficult to define and demarcate within words. MORPHEMIA attempts to bridge the gap between two radically different approaches to the task of the morphological segmentation of words: the time-consuming and expensive manual segmentation by linguists; and the completely unsupervised but theoretically blind automatic segmentation by language engineers, who try to bypass the problem by utilizing unsupervised machine-learning algorithms. The latter perform the task without the help of any a priori linguistic knowledge, extracting knowledge directly from the data (Siivola et al. 2007).

## Description

### A priori linguistic knowledge

MORPHEMIA uses two types of linguistic knowledge: lists of re-write rules for morphemes and a manually segmented sample of its lexical input.

MORPHEMIA uses five lists (L1 – L5) which contain re-write rules pertinent to five different types of morphemes and morpheme clusters that we assume to be relevant to the morphology of Modern Greek (MG). L1 contains “exceptions”, mainly rules for whole words whose morphological structure conforms to rare patterns, usually because of their ancient or foreign origin (these are often monosyllable and/or function words.) L2 contains rules for morphemes or morpheme clusters which can occur at the end of MG words and which may or may not coincide with the traditional “endings” listed in grammars of MG (e.g.  $\omega \rightarrow -\acute{\omega}$ ,  $\acute{o}\varsigma \rightarrow -\acute{o}\varsigma$ ,  $\epsilon\acute{\iota}\varsigma \rightarrow -\epsilon\acute{\iota}\varsigma$ ,  $\iota\kappa\acute{o}\varsigma \rightarrow -\iota\kappa-\acute{o}\varsigma$ ,  $\acute{\alpha}\delta\epsilon\varsigma \rightarrow -\acute{\alpha}\delta-\epsilon\varsigma$ , etc, the hyphens implying morpheme boundaries). L3 contains rules pertinent to traditional “roots” or clusters of them commonly occurring in MG compounds. (The latter can include derivational morphemes, e.g.  $\omicron\lambda\omicron\gamma \rightarrow -\omicron-\lambda\omicron\gamma-$ ,  $\omicron\iota\kappa\omicron\nu\omicron\mu \rightarrow -\omicron\iota\kappa-\omicron-\nu\omicron\mu-$ ,  $\iota\kappa\omicron\pi\omicron\acute{\iota}\eta\varsigma \rightarrow -\iota\kappa-\omicron-\pi\omicron\acute{\iota}\eta\varsigma-$ , etc.) L4 contains rules for (clusters of) derivational morphemes occurring to the right of “roots” (e.g.  $\alpha\tau\iota\sigma\acute{\mu}\acute{\epsilon}\nu \rightarrow -\alpha\tau-\iota\sigma-\acute{\mu}\acute{\epsilon}\nu-$ ,  $\epsilon\acute{\upsilon}\theta\eta\kappa \rightarrow -\epsilon\acute{\upsilon}-\theta\eta\kappa-$ ,  $\acute{\eta}\sigma\iota\mu \rightarrow -\acute{\eta}\sigma-\iota\mu-$ ,  $\acute{o}\tau\eta\tau \rightarrow -\acute{o}\tau-\eta\tau-$ , etc). Finally, L5 contains rules for (clusters of) derivational morphemes occurring to the left of “roots” and/or at the beginning of words ( $\alpha\nu\tau\iota \rightarrow -\alpha\nu\tau\iota-$ ,  $\xi\alpha\nu\alpha \rightarrow \xi\alpha\nu\alpha-$ ,  $\epsilon\pi\alpha\nu\alpha \rightarrow \epsilon\pi-\alpha\nu\alpha-$ ,  $\alpha\nu\alpha\delta\iota\alpha \rightarrow -\alpha\nu\alpha-\delta\iota\alpha-$ , etc).

Prior to its operation, MORPHEMIA needs to be supplied with a random and representative sample of the wordlist it will process. The words of the sample must be manually segmented into morphemes. The segmented sample will then function as MORPHEMIA’s “Golden Standard”. In essence, the manually segmented sample is assumed to be an implicit (yet adequate) statement of MG morphology. While creating the Golden Standard, the users of MORPHEMIA may introduce as much or as little linguistic knowledge as possible or desirable.

### Operation

The operation of MORPHEMIA is outlined in Figure 1. During its first iteration, MORPHEMIA uses the knowledge contained in L1 – L5 to segment the *sample*. This automatically segmented sample is compared against the Golden Standard. The metric  $b$  expresses the precision of the first segmentation:  $b = (\text{correctly segmented words} / N) \times 100$ , where  $N$  is the number of words comprising the sample.

Next, MORPHEMIA searches within the words of its entire *input* for the “target strings” contained in lists L1 – L5 (i.e. for the strings to the left of the rules’ arrows) and replaces them with the dummy character “\*”.

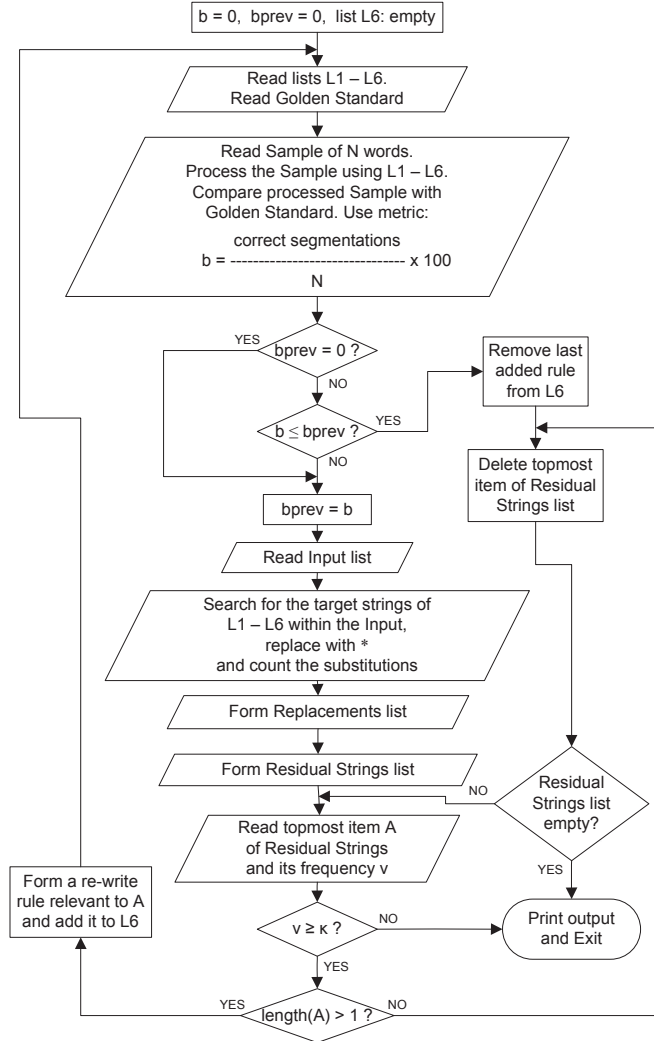


Figure 1. Flowchart of the MORPHEMIA algorithm.

Following that, MORPHEMIA counts the frequency of the substrings which were not replaced by “\*” and creates the Residual Strings list.

The most frequent of these sub-word units, “A”, is then tested for its validity as a legitimate MG morpheme. A relevant rule ( $A \rightarrow -A-$ ) is formed and added to the as yet empty list L6. Then, MORPHEMIA re-segments the sample using the contents of L1 – L5 but also of L6. If the new  $b$  metric is

smaller than or equal to the previous one, the rule is removed from L6, the relevant string is removed from the Residual Strings list and the algorithm repeats the process with the next most frequent residual string. However, if the new  $b$  is greater than the previous one, the string is accepted as a legitimate morpheme and the relevant rule remains in L6, thus becoming part of the a priori knowledge to be used in the next iteration. The user may define a floor value  $\kappa$  as the lowest frequency which determines if a substring will be considered at all (in our experiments,  $\kappa = 1$ ). Note that, although one-character long morphemes can be justified in MG morphology, MORPHEMIA disallows them to avoid over-segmentation of the input.

MORPHEMIA continues its operation until the Residual Strings list is empty. The contents of L6 represent the new morphological knowledge extracted by the algorithm.

### Brief discussion of experimental results

A number of experiments have been conducted, using an input list of 226,857 different words, a Golden Standard (GS) of N=1000, and a Final Evaluation Sample (FES) of N=1000, different from GS. For lack of space here, we report on MORPHEMIA's worst and best performances. **Worst:** a priori knowledge: 69 rules (L1: 0, L2: 39, L3: 6, L4: 7, L5: 17),  $b_{GS-initial}$ : 11.70%,  $b_{GS-final}$ : 41.70%, input characters affected: 56.53%, L6: 185 rules,  $b_{FES-final}$ : 25.20%. **Best:** a priori knowledge: 2527 rules (L1: 22, L2: 508, L3: 1552, L4: 258, L5: 187),  $b_{GS-initial}$ : 53.70%,  $b_{GS-final}$ : 74.50%, input characters affected: 75.22%, L6: 127 rules,  $b_{FES-final}$ : 60.10%.

As expected, MORPHEMIA's performance depends on the amount of a priori knowledge, especially on the size of L3 (rules for "roots"). Interestingly, the candidate morphemes *rejected* by the  $b$  metric evaluation process are a valuable source of new candidate roots. These can be manually selected and added to the knowledge included in L3. Thus, MORPHEMIA can also prove a valuable tool for building morphological lexicons: the ratio "input words / strings considered" is roughly 11/1.

### References

- Kirchhoff, K. and Sarikaya, R. 2007. Processing morphologically rich languages. Research Tutorial at the 8th Annual Conference of the International Speech Communication Association, Interspeech 2007. Antwerp, Belgium.
- Siivola, V., Creutz, M. and Kurimo, M. 2007. Morfessor and VariKN machine learning tools for speech and language Technology. Proceedings of the 8th Annual Conference of the International Speech Communication Association, Interspeech 2007. Antwerp, Belgium.

# The acquisition of temporal categorical perception by Japanese second language learners

Naoko Kinoshita

Integrated Education Center, Meikai University, Japan

## Abstract

This study reports research into the development of perception of special mora categories by Korean learners of Japanese as a second language. First, a test which measured boundary width and the boundary point of distinction between singletons and special mora was used to establish participants' categorical perception across four levels of proficiency. Then a two year longitudinal study followed the development of 14 learners, during which time the same test was administered three times. The results demonstrated: 1) that the special mora perceptual categories of second language learners developed in some areas, but not in others and 2) that there was a categorization of the special mora over time.

Key words: categorical perception, rhythm, language acquisition, mora timing

## Introduction

Japanese is a mora-timed language in which each mora receives roughly the same articulation time. There are three special mora (long vowels, geminate consonants, and nasal consonants) perceived to receive twice as much articulation time. For example, while *obasan* means 'aunt' in Japanese, *oba:sa:n* means 'grandmother'. This distinction is difficult for learners of Japanese to acquire. As yet, there has been little research which examines the acquisition of the mora-timed rhythm. (However, see Kinoshita 2007, Nishogori *et al.* 2002, Toda 1998, and Uchida 1998)

This study attempts to determine the process of acquisition of categorical perception of special mora by Korean learners of Japanese. It approaches the problem using both cross-sectional and longitudinal methodologies.

## Research Method

### Participants

The participants were 21 Korean university students who were studying Japanese at a university in Busan. Based on the Oral Proficiency Interview (OPI) held at the beginning of the study, these included 6 advanced learners, 9 intermediate and 6 beginning learners. Of these 21, 14 learners participated in the longitudinal study. In addition to the Korean learners of Japanese, 7 Japanese native speakers of the Tokyo dialect (JS) were involved in the study. All participants were compensated for their time.

### **Instruments and Procedure**

In order to measure the participants' perception of Japanese mora rhythm, a perceptual experiment was designed. Eight minimal pairs were recorded by a female native speaker of the Tokyo dialect in meaningful carrier sentences. Two sentences contrasted the long/short vowel distinction (R), two contrasted the nasal /N/ distinction (N) and four contrasted the geminate/singleton consonant distinction (Q). The selection of the minimal pairs was also controlled for accent (HL and LH) and, in the case of the geminate/singleton, consonant (fricative and plosive) (see Toda 1998).

Fifteen samples of each minimal pair were then generated from the natural recording using acoustic analysis software (*OnseiKoubou* NTT Advanced Technology). The vowels, nasals and consonants were lengthened at increments of 10% from -20% to +120% with 0% being the length of the short sound, and 100% being the length of the long sound.

These samples were presented to the participants on an audio CD, twice, and in random order. They were required to indicate if the stimulus word in each sentence was long or short by circling the word. This resulted in 240 judgements per participant. The entire procedure took about 11 minutes per CD. For the 14 longitudinal study participants, the data was collected twice more at yearly intervals between collections.

### **Analysis**

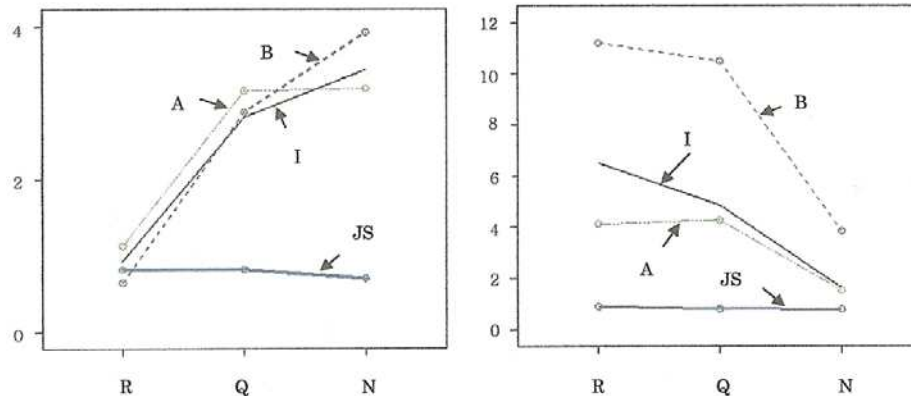
The distinction between the temporal categories was operationalized by boundary point (BP) and boundary width (BW). BP is the point of time at which the perception of whether the sound is long or short changes. It is recorded at the point where the participants are 50% certain that the sound is long. BW is the temporal width. This is the time taken between participants' 20% certainty that the sound is long to their 80% certainty of it. BP and BW were measured for each participant and each sentence.

In order to respond to the research questions, four repeated measures ANOVAs were undertaken. The first two compared level (NS, beginner, intermediate and advanced) with rhythm type (R, N and Q) for both BP and BW. Similarly, the second two compared the time (year 1, year 2, and year 3) with the rhythm types (R, N, and Q). Significance was set at 5%.

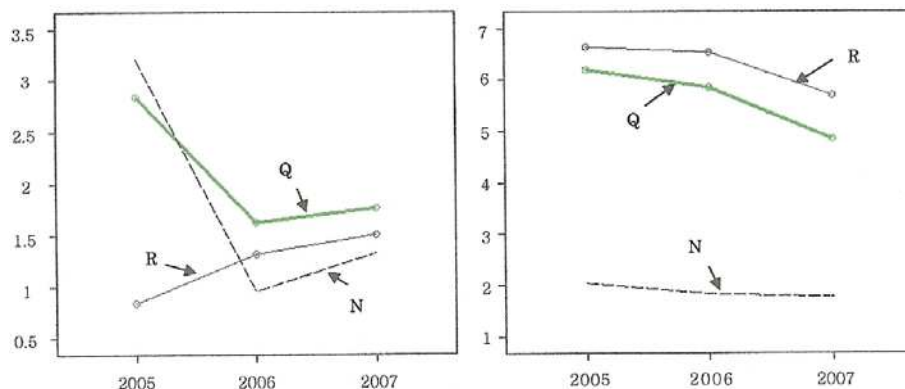
### **Results**

First, there were no significant differences between the four proficiency levels for R. However, the Korean groups did have boundary points significantly displaced from those of Japanese native speakers for N and Q. There were no differences between Korean learners' proficiency levels. However, there was a proficiency level difference for Boundary Width. The

beginners perceived larger BWs for all three rhythm types than the intermediate and advanced groups. There were no observed differences between the other learner groups and JS. These results are represented visually in figures 1 and 2.



Figures 1 and 2. The BP and BW perception of R, Q and N (B = beginning learners, I = intermediate learners, A = advanced learners, JS = Japanese native speakers).



Figures 3 and 4. The change in BP and BW perception of R, Q and N over time.

The BP longitudinal data demonstrated a trend toward a general rhythmic perceptual category. Initially, in 2005, R was very close to the JS score. However, by 2006 it had significantly increased from its 2005 value, away from JS perception. In contrast, the BPs of Q and N become significantly closer to that of the JS from 2005 and 2006. In both 2006 and 2007 there was no significant difference between the R, Q or N. There was no change for BW over the three times it was measured in the cross sectional study and the



significantly narrower BW demonstrated in the longitudinal data was maintained (see figures 3 and 4).

### Discussion and Conclusion

The analyses of the results have provided us with several points of discussion. First, there does appear to be some developmental change in the perception of rhythm. This is evident in both the BW data in the cross-sectional results and the BP data in the longitudinal study. These results mirror those of Kinoshita (2007) and Nishigori *et al.* (2002) However, even over the two year duration of this study, there were areas of perception which did not change, indicating that these aspects of rhythm may need to be taught to learners.

Perhaps the most interesting result was that the Korean learners' boundary point timing showed they had 'categorized' (see Ryalls 2003, p. 44) the special mora. Initially the BPs were different for the three special mora, however, in the second and third year of the study, they moved closer together, with R even moving away from the JS norm.

### Acknowledgements

This work was funded in part by the Grant-in-Aid for Scientific Research (C), "The acquisition of Japanese special mora by second language learners", no. 19520463 from JSPS.

### References

- Kinoshita, N. 2007. The Acquisition Processes of the Perception of Japanese Long and Short Vowels: The Case of Korean Learners of Japanese as a Second Language, *The Korean Journal of Japanology*. 72, 1-12.
- Nishigori, J., Hwan, Y., Park, Y. 2002. Kankokujin gakusyusya no Nihongo sokuon no chikaku ni kansuru kenkyu: gakusyu reberubetsu tokusei to bogo niyoru setsumeimei no kouka, *Nihongo Kenkyu*22, 103-118.
- Ryalls, J. 1996. *A Basic Introduction to Speech Perception*, Tokyo, Kaibundo.
- Toda, T. 1998. Perceptual categorization of the durational contrasts by Japanese learners. *Studies in Language and Literature* 33, University of Tsukuba, 65-82.
- Uchida, T. 1998. Categorical Perception of Relatively Steady-Static Speech Duration in Japanese Moraic Phonemes, *Journal of Phonetic Society of Japan* 2-3, 71-86.

# **Tongue movements and syllable onset complexity: ultrasound study**

Tanja Kocjancic

Speech and Hearing Sciences, Queen Margaret University, UK, and  
Centre for Speech Technology Research, University of Edinburgh, UK

## **Abstract**

In this study ultrasound was used to investigate tongue movements in syllables with different number and type of onset consonants. Ultrasound recordings provided the information of the distance the tongue travels over a target, and audio recordings of the time needed. The speed of the tongue's travel was calculated from the two measurements. Results of ten speakers have shown that both duration and distance travelled increase with an increased number of onset segments, but that distance travelled is additionally influenced by the type of the segment, as is speed. Duration also seemed to be the least speaker-dependant of the three parameters.

Key words: tongue movement, ultrasound, syllable onset

## **Introduction**

Ultrasound is a safe and non-invasive method which enables visualisation of tongue movement by depicting the upper tongue surface (Stone 2005). To obtain an ultrasound image of the tongue, the probe is placed under the speaker's chin and the inside of the mouth is scanned by emitting high frequency ultrasound waves. These waves are reflected at a boundary between mediums of different density ( tongue/air and tongue/bone) and detected by the probe. After the detection the point of reflection is calculated and an image is created at that point.

Characteristics of differently structured syllable onsets have been investigated in several studies (Greenberg et al. 2003; Crystal and House 1990). The findings generally suggest that syllable duration is mainly affected by the syllable's stress and the number of segments. Syllable duration increases with an increasing number of segments in both stressed and unstressed condition. Stressed syllables and their segments have greater duration than their unstressed counterparts.

The aim of this study was to investigate how tongue movements are affected by changing the type and number of syllable onset segments. Tongue movements are described by the distance travelled by the tongue over an utterance, the duration of an utterance, and the tongue speed. It was hypothesized that both measurements will increase with the increasing number of onset segments, and that speed will vary, depending on the type of segments.

## **Methodology**

Ultrasound data of ten native English female speakers, aged between 19 and 30 years, was analysed in this study. Speech material consisted of six mono-syllabic real English words: pay, say, lay, play, slay, splay.

Midsagittal view of the tongue was recorded with Concept M6 (Dynamic Imaging) ultrasound with a frame rate of 30fps, and a special helmet was used to fix the probe under the speaker's chin. Participants repeated each of the words five times in a frame sentence "a [word] today". Both ultrasound and audio signals were recorded at the same time using Articulate Assistant Advanced, which allows temporal synchronisation of the two signals.

Articulate Assistant Advanced was used for annotating and tracing tongue contours on the recorded ultrasound images. All the reported results are for the "a [word]" part of the recording. The travelled distance of a target utterance was calculated as the sum of average nearest neighbour distances (aNND) between every pair of consecutive tongue contours of the utterance. aNND is an average of all the nearest neighbour distances measured between the points on the two contours of a pair. Duration was measured from audio signal, and the tongue speed was calculated as the distance travelled over duration, to give information about the relationship between the two measurements.

## **Results**

### **Duration**

The results showed that duration increases with the increasing number of syllable segments (Figure 1a) and that it is statistically significant between all pairs of targets (Figure 1d, solid lines). Additionally, most of the speakers showed the same pattern of increasing duration: targets with single onsets shorter than those with two consonants clusters, which were also shorter than the three consonants one.

### **Distance travelled by the tongue**

Overall, the distance travelled by the tongue did increase with the increased number of onset segments (Figure 1b) but the increase was not stat. sig. between all of the syllable pairs (Figure 1d, dashed lines). Distribution of the measurements was also more variable than in the case of duration, and individual speakers showed less similar patterns of increasing the distance travelled by the tongue over a syllable. Not all speakers had greater measurements for targets with clustered onset than for those with single onset.

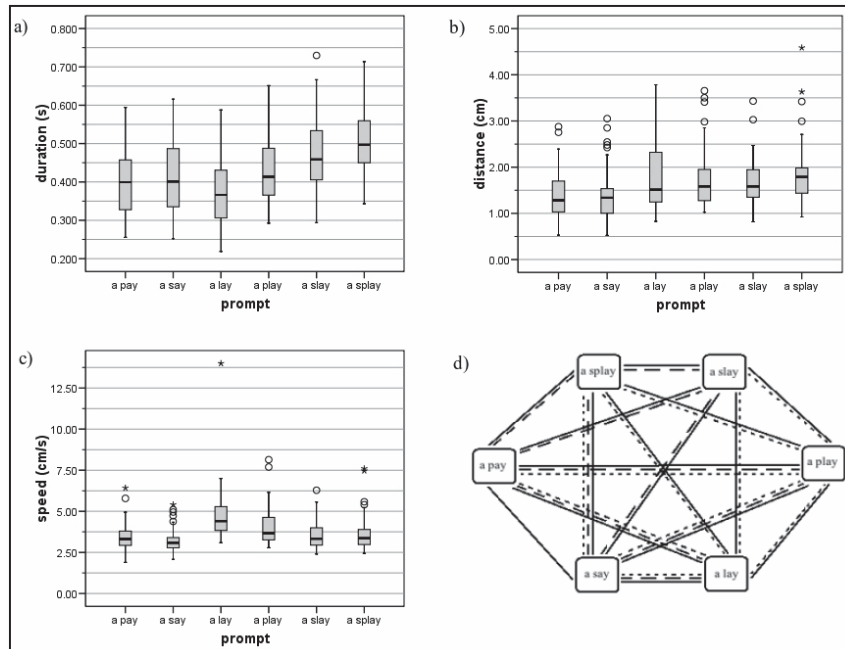


Figure 1. Duration (a), tongue's distance travelled (b), tongue speed (c) and stat. sig. (Wilcoxon Signed Rank Test,  $p < 0.05$ ) between pairs of targets (d; dashed line = distance travelled, solid line = duration, dotted line = tongue speed.).

### Speed of the tongue's travelling

Speed of the tongue's travelling is presented in Figure 1c. The tongue travelled the fastest over "a lay" and the slowest over "a say" with most of the target pairs being stat. sig. different from each other (Figure 1d, dotted lines). Number of onset segments did not influence tongue speed, and individual speakers showed very different patterns of increasing tongue speed over targets, although eight out of ten speakers had the fastest tongue speed over "a lay".

### Discussion and conclusion

Analysis has shown that duration increases with the addition of segments to syllable onset and thus confirmed the results of previous studies (Greenberg et al. 2003; Crystal and House 1990). Results of distance travelled measurements also showed the influence of the number of onset segments, as single onset targets were shorter than clustered ones. However, they were also influenced by the type of the segment. As expected, targets with single onset /p/ and /s/ were shorter than those with single /l/, which had a distance

travelled more similar to the targets with clustered onsets. /p/ is not a lingual consonant and does not contribute any tongue movement to the distance travelled, and /s/ has less tongue movement than /l/. In contrast, /p/ does contribute to the distance travelled in case of clustered onsets as it seems to make movement for /l/ more prominent and less restricted than in case of /sl/ clusters. Tongue speed, on the other hand, does not seem to depend on the number of segments, but mainly on the type.

Based on the results of this study it can also be concluded that duration of spoken target is less speaker-dependant than distance travelled. The latter depends on the size of individual speech organs and can not be adapted as duration can. Consequently, tongue speed is the most speaker-dependant and is appropriately adapted depending on the demands of the space over which the tongue has to travel, and the demands of appropriate speech timing.

This study has demonstrated that ultrasound is sensitive enough to describe continuous tongue movement during speech although it has some limitations. The most important is that it does not produce an image of raised tongue tip with air pocket below it, and thus we miss information about that part of the tongue. Measurements could be also affected by the probe which could potentially restrict jaw movement. Mooshammer et al. (2003) have shown that jaw position is low for /l/ and high for /s/. However, this effect is expected to affect all participants.

### **Acknowledgements**

Many thanks to my supervisor, Dr Nigel Hewlett. Financial support provided by the Marie Curie EdSST programme.

### **References**

- Articulate Instruments Ltd. 2007. Articulate Assistant Advanced User Guide: Version 2.07. Edinburgh, UK: Articulate Instruments Ltd.
- Crystal, T. H. and House, A. S. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88, 101-112.
- Greenberg, S., Carvey, H., Hitchcock, L. and Chang, S. Y. 2003. Temporal properties of spontaneous speech - a syllable-centric perspective. *Journal of Phonetics* 31, 465-485.
- Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P. and Fuchs, S. 2003. Coordination of lingual and mandibular gestures for different manners of articulation. In *Proceedings of the 15th ICPhS*, 81-84, Barcelona, Spain.
- Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics* 19, 455-501.

# **The prosodic and nonverbal deficiencies of French- and Finnish-speaking persons with Asperger Syndrome**

Mari Lehtinen

Department of Romance Languages, University of Helsinki, Finland

## **Abstract**

This paper briefly presents a research project directed towards the prosodic and nonverbal deficiencies of French- and Finnish-speaking persons with Asperger Syndrome (AS). AS is a form of high-functioning autism and it is characterized e.g. by qualitative impairment in social interaction and by stereotyped and restricted patterns of activities and interests. Though persons afflicted with AS generally have no significant delay in language acquisition, and their speech typically lacks significant abnormalities, the language use and the prosody of persons with AS is often atypical: people with AS often have a limited range of intonation, their speech may be overly fast, jerky or loud, and they also typically have abnormal nonverbal behaviours (such as eye contact, facial expressions, postures and gestures).

## **Introduction**

This paper aims at providing a brief presentation of a post-doctoral research project that will be launched in the beginning of 2009. The project is directed towards the prosodic and nonverbal deficiencies of French- and Finnish-speaking persons afflicted with Asperger Syndrome (AS). AS is considered as a form of high-functioning autism (HFA) and it is characterized e.g. by qualitative impairment in social interaction and by stereotyped and restricted patterns of activities and interests (APA 2000). Though people with AS generally have no significant delay in language acquisition (APA 2000), and their speech typically lacks significant abnormalities, their language use and the prosody of speech are often atypical (Klin 2006): people with AS often have a limited range of intonation, their speech may be overly fast, jerky or loud. According to Partland and Klin (2006), people afflicted with AS also typically have abnormal nonverbal behaviours (such as eye contact, facial expressions, postures and gestures).

So far, the speech prosody and the nonverbal behaviors of people with As have been studied mainly in the fields of psychology, neurology and paediatrics. The project presented here aims at providing a linguistic study (including a phonetic perspective) contrasting the prosodic and nonverbal features of French- and Finnish-speaking people diagnosed with AS. In addition to the general contribution that a linguistic (and a phonetic) approach may bring to the field of AS, the fact of contrasting two unrelated and fundamentally different languages such as French and Finnish can be

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.

expected to produce interesting results concerning the role of language- and culture-specific features in the development of speech and nonverbal behaviour of people with AS.

The data will consist of audiovisual material recorded from discussions where under 12-year-old French- and Finnish-speaking children are being diagnosed by a paediatrician specialised in AS.

Methodologically, the study will be based on conversation analysis (Sacks, Schegloff and Jefferson 1974; Hakulinen 1997; Sorjonen *et al.* 2006; Peräkylä 2004) and, more precisely, on the interactional approach to the study of prosody (Couper-Kuhlen and Selting 1996). The contextualization theory of Gumperz (1992) will also have a central role in the project: indeed, the prosodic and nonverbal features constituting the object of the study will be considered as “contextualization cues” in the sense of Gumperz.

## **Background**

So far, the prosody of persons with AS has been studied mainly within the fields of psychology and neurology with the help of mismatch negativity (MMN), which is a component of the event-related potential (ERP), arising from the electrical activity in the brain (Korpilahti *et al.* 2007; Kujala *et al.* 2005). The results of the MMN studies suggest that people with AS have problems to recognise affective prosody. It has also been shown that persons with AS have difficulties to produce affective prosodic patterns (Scott 1985). The speech of people with AS is often melodically monotonous, and it may also be overly fast, jerky or loud (Partland and Klin 2006). People afflicted with AS often have atypical nonverbal behaviours – such as eye contact, facial expressions, postures and gestures (Partland and Klin 2006) – and they may also have difficulties to recognise nonverbal signs of the interlocutor (Scott 1985). It has also been shown that it is more difficult for children afflicted with AS than for other children to use the linguistic context for the interpretation of a verbal message (Loukusa *et al.* 2007).

## **Objectives**

This project is directed towards comparing the prosodic and nonverbal features characterising the interaction of French- and Finnish-speaking people with AS. The objective of the study is to find out in what measure the prosodic and nonverbal deficiencies of persons afflicted with AS depend on purely diagnostic factors, and what is the role of language- and culture-specific features in the manifestations of these features.

Choosing to contrast precisely French and Finnish is justified by at least four factors: 1) French and Finnish are unrelated languages (Indo-European vs. Uralic language); 2) there is no close contact between these languages; 3)



they are spoken in rather different cultural environments; 4) the prosodic structures of these languages are completely different.

In addition to comparing the two languages, one of the goals of the project is to acquire new information concerning the connection (the cooperation / the lack of cooperation) between the prosodic and nonverbal means of communication used by people afflicted with AS.

## Data

The data will consist of audiovisual (DVD) recordings from five French and five Finnish diagnostic discussions, the diagnostic result of which is positive. The informants will be 4–11-year old native French- and Finnish-speaking children who are being diagnosed by a paediatrician specialised in AS. Choosing children as informants is particularly interesting for the following reasons: 1) AS is typically diagnosed in 4–11-year-old children (Partland and Klin 2006); 2) the influence of experience and practise is less important in the case of children than in the case of adults.

As being five times more common for males than for females, most informants will be male. If possible, persons having been diagnosed with other neurological deviations (such as ADHD, AD, HD, Tourette syndrome or OCD) and/or being afflicted with mood disorders requiring medication (such as depression, anxiety or bipolar disorder) at the moment of the recorded discussion are preferably not chosen as informants, because the aforesaid states may affect the prosodic and nonverbal means of communication used by the person.

## References

- American Psychiatric Association (APA) 2000. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edition, Text Revision, DSM-IV-TR. Arlington, American Psychiatric Association.
- Couper-Kuhlen, E. and Selting, M. 1996. Towards an interactional perspective. In Couper-Kuhlen, E. and Selting, M. (eds.) 1996, *Prosody in conversation*. Interactional studies, 11-56. Cambridge, Cambridge University Press.
- Danon-Boileau, L. and Morel, M.-A. 2007. Approche linguistique du discours autistique : quelques remarques. In Touati, B., Joly, F. and Laznik, M.-C. (eds.) 2007, *Langage, voix et parole dans l'autisme*, 335-340. Paris, PUF.
- Di Cristo, A. 1998. Intonation in French. In Hirst, D. and Di Cristo, A. (eds.) 1998, *Intonation Systems. A Survey of Twenty Languages*, 195-218. Cambridge, Cambridge University Press.
- Gumperz, J. J. 1992. Contextualization Revisited. In Auer, P. and Di Luzio, A. (eds.) 1992, *The Contextualization of Language*, 39-53. Amsterdam/Philadelphia, John Benjamins.
- Hakulinen, A. 1997. Vuorottelujäsennys. In Tainio, L. (ed.) 1997, *Keskusteluanalyysin perusteet*, 32-55. Tampere, Vastapaino.



- Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytty, S. Pauls, D.L. and Moilanen, I. 2007. Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders* 37(8), 1539-1549.
- Kujala, T., Lepistö, T., Nieminen-von Wendt, T. Näätänen, P. and Näätänen, R. 2005. Neurophysiological evidence for cortical discrimination impairment of prosody in Asperger syndrome. *Neuroscience Letters* 383(3), 260-265.
- Loukusa, S., Leinonen, E., Kuusikko, S., Jussila, K., Mattila, M.-L., Ryder, N., Ebeling, H. and Moilanen, I. 2007. Use of context in pragmatic language comprehension by children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders* 37(6), 1049-1059.
- Morel, M.-A. and Danon-Boileau, L. 1998. *Grammaire de l'intonation. L'exemple du français oral*. Paris, Ophrys.
- Nieminen-vonWendt, T. 2004. On the origins and diagnosis of Asperger syndrome. A clinical, neuroimaging and genetic study. Academic dissertation. University of Helsinki. Helsinki, Yliopistopaino.
- Partland, J. and Klin, A. 2006. Asperger's syndrome. *Adolescent Medicine Clinics* 17(3), 771-88.
- Peräkylä, A. 2004. Potilaan rooli psykoanalyysissä ja yleislääkärin vastaanotolla. In Alapuro, R. and Arminen, I. (eds.) 2004, *Vertailevan tutkimuksen ulottuvuuksia*, 245-258. Helsinki, WSOY.
- Sacks, H., Schegloff, E. and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696-735.
- Scott.D.-W. 1985. Asperger's syndrome and non-verbal communication: a pilot study. *Psychological Medicine* 15(3), 683-687.
- Sorjonen, M.-L., Raevaara L., Haakana, M., Tammi, T. and Peräkylä, A. 2006. Lifestyle discussions in medical interviews. In Heritage, J. and Maynard, D. (eds.) 2006, *Communication in Medical Care. Interaction between primary care physicians and patients*, 340-378. Cambridge, Cambridge University Press.

# **The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels**

Angelos Lengeris

Department of Phonetics and Linguistics, University College London, UK

## **Abstract**

This study investigated the effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. The trainees identified and produced English bVt words before and after receiving five sessions of High Variability Phonetic Training (multiple speakers in multiple contexts). All of the trainees improved in their overall identification of English vowels. A perception experiment showed that their post-training productions were more accurately identified by native English speakers than their pre-training productions. Taken together, the results demonstrate that speakers of a language with a simple 5-vowel system (viz. Greek) can improve in both perceiving and producing the vowels of a language with a complex system (viz. English) via intensive laboratory training.

Key words: auditory training, second-language learning

## **Introduction**

Second language (L2) learners are usually found to experience great difficulty in perceiving and producing vowels in the target language. When both members of an L2 vowel contrast assimilate to the same first language (L1) vowel category, learning is more challenging than cases where each member assimilates to a different L1 vowel category (Best 1995, Flege 1995).

L2 learners' perception of novel contrasts can improve via auditory training (with most studies focusing on consonant learning, e.g. Logan et al. 1991) with gains in perception transferring in production both for consonants (Bradlow et al. 1997) and vowels (Lambacher et al. 2005). However, research on vowel training has exclusively examined Japanese L1 speakers (e.g. Lambacher et al. 2005, Nishi and Kewley-Port 2007) with the exception of Iverson and Evans (2007) who trained German and Spanish L1 speakers' perception of English vowels but did not examine gains in production.

The present study examined whether native speakers of Greek (Gr) can be trained both in perceiving and producing the Southern British English (SBE) vowels. Gr speakers do not have tense-lax or long-short distinctions in L1 and consequently find most SBE vowels difficult to perceive especially when durational information is reduced (Lengeris and Hazan, 2007).

## Method

### Participants

Ten Athenian Gr speakers (mean age = 23 years, range = 18-35 years) participated in the training programme (twenty speakers were tested in total but only the results for the first ten are reported here). Another ten Athenian Gr speakers (mean age = 26 years, range = 18-42 years) served as controls, i.e. received no training. All of the participants had 10-12 years of formal English instruction in Greece with none having spent a period of more than one month in an English-speaking country. They all passed a pure-tone hearing screening at frequencies from 250 to 4000 Hz at 20 dB SPL.

### Procedure

Following Logan et al. (1991), the training programme consisted of a pre-test phase, a training phase, and a post-test phase. In pre- and post-test both groups were given a large battery of perceptual tests, however, this study reports only on their identification of SBE vowels. The stimuli consisted of ten SBE bVt words (all SBE monophthongs except /ʊ/ in a ten-alternative forced-choice task) spoken by two SBE speakers (1 m, 1 f). The post-test also included a *generalization* test where a new SBE speaker (f) was added. Participants responded to 40 trials in the pre-test (2 speakers  $\times$  10 vowels  $\times$  2 times) and 60 trials in the post-test (3 speakers  $\times$  10 vowels  $\times$  2 times).

After completing the pre-training identification task, the participants were asked to read aloud from a screen the 10 bVt words they had previously attempted to identify (each word was read two times). The participants had therefore encountered the target words before producing them although the task was not a direct repetition one. Two SBE speakers, aged 26-28 years, identified Gr speakers' pre- and post-training SBE vowel productions in a 10AFC task (all productions were fully randomized).

The training stimuli and procedure were the same as in Iverson and Evans (2007). Five SBE speakers (2 m, 3 f) recorded 140 target words containing 14 SBE vowels, arranged in 4 groups (/i: ɪ aɪ eɪ/, /u: əʊ ɜ:/, /ɒ əʊ ɔ:/, /e æ a:/), i.e. each group consisted of 10 minimal pairs. The trainees completed 5 sessions of computer-based auditory training with feedback within 2-3 weeks with a different speaker each day. In each session the participants responded to 225 trials (in 45-60 minutes); they heard an English word and chose one of 3 or 4 candidates as displayed on a computer screen. If a correct answer was given "Correct!" was displayed on the screen, a cash register was heard and the target word was repeated once. If an incorrect answer was given "Wrong" was displayed on the screen, two beeps were heard and both the target and the (wrongly) chosen word were repeated twice. Percent correct identification was displayed at the end of each session.

## Results

### Auditory training and L2 perception

Figure 1 (left panel) shows percent correct pre- and post-training identification of SBE vowels for the trained and the control group. A two-way ANOVA with Group (trained, control) and Test (pre, post, generalization) as factors showed significant main effects of Group [ $F(1,10) = 7.3$ ,  $p = 0.015$ ] and Test [ $F(2,36) = 13.9$ ,  $p < 0.001$ ] and a significant Group  $\times$  Test interaction [ $F(2,36) = 10.6$ ,  $p < 0.001$ ]. The trained group improved significantly (Bonferroni adjusted  $p < 0.05$ ) from pre- ( $M = 57\%$ ) to post- and generalization test ( $M = 78\%$  and  $81\%$  respectively) whereas the control group did not improve from pre- ( $M = 56\%$ ) to neither post- nor generalization test ( $M = 58\%$  and  $57\%$  respectively).

### Auditory training and L2 production

Figure 1 (right panel) shows percent correct identification of trained and control groups' pre- and post-training SBE vowel productions as judged by the SBE L1 speakers. A two-way ANOVA with Group (trained, control) and Test (pre, post) as factors showed a significant main effect of Test [ $F(1,18) = 17.6$ ,  $p < 0.001$ ] and a significant Group  $\times$  Test interaction [ $F(1,18) = 6.6$ ,  $p = 0.018$ ]. The trained group improved significantly (Bonferroni adjusted  $p < 0.05$ ) from pre- ( $M = 53\%$ ) to post-test ( $M = 70\%$ ) whereas the control group did not improve from pre- ( $M = 57\%$ ) to post-test ( $M = 60\%$ ).

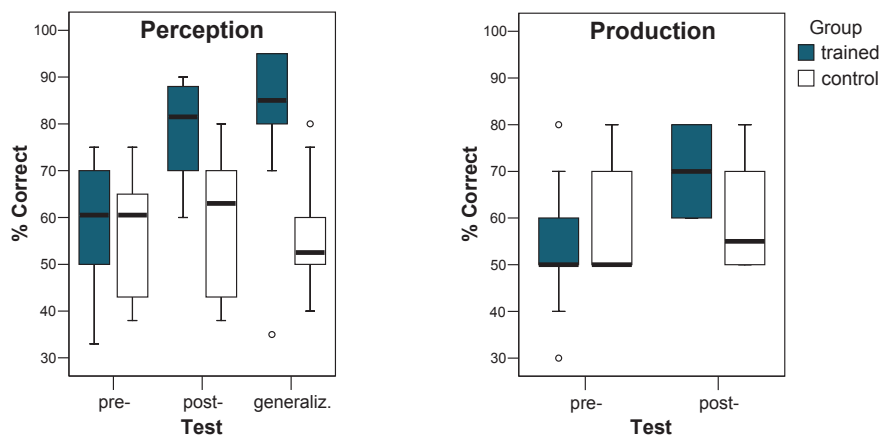


Figure 1. Percent correct identification of SBE vowels in pre- post- and generalization tests for trained and control group (left panel) and percent correct identification of trained and control groups' pre- and post-training SBE vowel productions as judged by SBE speakers (right panel).

## Discussion

This study investigated the effectiveness of auditory phonetic training on Gr L1 speakers' perception and production of SBE vowels. It was found that intensive auditory training improved the trainees' identification and production of L2 vowels even though they did not receive any explicit production training. This finding supports not only the notion of plasticity in L2 learning but also a link between perception and production when acquiring an L2.

## Acknowledgements

I am grateful to Valerie Hazan for her useful comments on an earlier version of this paper and to Paul Iverson for kindly providing me with the training materials and software. This study was supported by a research grant from the A.G. Leventis Foundation.

## References

- Best, C.T. 1995. A direct realist view of cross-language speech perception. In Strange W. (eds.) 1995, *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Timonium, MD: York Press.
- Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., and Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101, 2299–2310.
- Flege, J.E. 1995. Second language speech learning theory, findings, and problems. In Strange W. (eds.) 1995, *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. Timonium, MD: York Press
- Iverson, P. and Evans, B. 2007. Auditory training of English vowels for first-language speakers of Spanish and German. *Proc. of the 16th Intern. Congress of Phonetic Sciences*, 1625–1628, Saarbrücken, Germany.
- Lambacher, S.G., Martens, W.L., Kakehi, K., Marasinghe, C.A. and Molholt, G. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics* 26, 227–247.
- Lengeris, A. and Hazan, V. 2007. Cross-language perceptual assimilation and discrimination of Southern British English vowels by native speakers of Greek and Japanese. *Proc. of the 16th Intern. Congress of Phonetic Sciences*, 1641–1644, Saarbrücken, Germany.
- Logan, J.S., Lively, S.E., and Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89, 874–886.
- Nishi, K. and Kewley-Port, D. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research* 50, 1496–1509.

# **Phonetic convergence and language talent within native-nonnative interactions**

Natalie Lewandowski, Travis Wade and Grzegorz Dogil

Institute for Natural Language Processing, University of Stuttgart, Germany

## **Abstract**

The notion of phonetic convergence covers all adaptations in articulatory and acoustic features towards those of a communicative partner, or in other terms an increase in segmental and suprasegmental similarity between them (Pardo 2006). Up until now most of the experiments on convergence were designed for monolingual dyads, with very few investigations of convergence in native-nonnative interactions in a foreign language learning environment. We tried to analyze the convergent behavior of nonnative speakers of English in dialog with native speakers and the persistence of the effect in relation to their rated phonetic talent. In this paper we present first results for a global measurement of convergence - the comparison of amplitude envelope signals.

Key words: phonetic convergence, talent, foreign language, native-nonnative interaction

## **Introduction**

Accommodating strategies (such as convergence, divergence, maintenance and complementarity) may be used to achieve solidarity with or dissociation from a partner, in a dynamic setting with online feedback (Giles et al. 1991). As Wedel (submitted) points out, convergence and divergence can also be part of a simultaneous process, with some features underlying an automatically driven positive accommodation and some functioning as identity markers for group membership and therefore subject to divergence. In order to account for such apparently countervailing and still simultaneously observed strategies, they need to be embedded in a dynamical systems framework of language that accounts amongst others for more fine-grained usage-induced changes.

Thus in this paper we assume a usage-based account that allows for the storage and processing of fine phonetic detail and additional social information - an exemplar theoretical account (e.g. Johnson 1997, Pierrehumbert 2001). It has been suggested that normalization processes are not needed in an exemplar model of perception because “the model retains the variability encountered in speech [and] it is able to cope with the variability that it encounters in new tokens” (Johnson 1997:162). An exemplar-based account could also provide an explanation for social accommodation processes, since more recent and more frequently heard

exemplars could guide the typical productions within a speech community and therefore lead to the adaptation of the prominent speech patterns. Although exemplar-based production models also provide a straightforward explanation for the nature of the perception-production loop, with both processes relying on the same pool of exemplars (e.g. Pierrehumbert 2001), this does not imply that a phonetic target is necessarily realized as a perfect match (e.g. due to noise in the motor control and execution).

Since pronunciation seems to have a separate and special status within second language acquisition, we hypothesize that for pronunciation convergence towards a foreign language speaking partner, be it conscious or not, not only a certain proficiency in that language is crucial but also some degree of phonetic talent.

## **Method**

All participants of the present study have also been tested for their phonetic language talent in another ongoing project (see Jilka et. al 2008 for details). Their pronunciation talent based on their performance in various perception and production tests has been rated on a scale from 1 to 6, 1.0 standing for exceptionally talented and 6.0 for absolutely not talented.

## **Experimental design**

The participants were 30 native speakers of German that were involved in two dialogs with native speakers of General American (male) and Southern Standard British English (female). The elicitation technique used was a Diapix-task (Bradlow et al. 2007), a picture matching game in which participants had to identify ten differences between their pictures without seeing the partner's picture. The control task consisted of reading a list of words that contained words from the two picture-sets and unrelated filler words before and after each Diapix-dialog. The whole experimental session thus consisted of the following steps: first reading of the word list, Diapix-dialog A with the GA speaker, second reading of the word list, Diapix-dialog B with the SSBE speaker, third reading of the word list.

## **Analysis**

Participants' and native speakers' word-level productions were transformed to amplitude envelope signals over 4 log-spaced frequency ranges, equalized for total amplitude, and compared by cross-correlation to estimate a match. Match values range from 0 to 1, with 1 indicating a perfect match and 0 no match. Since these signals do not explicitly include any specific phonetic cues such as formant frequencies but are a (relatively) compact and transparent representation of information that is present in various forms

throughout the auditory system, the match value provides a rough global measurement of spectrotemporal similarity (Wade et al. submitted).

## Results

We first compared mean match values from a test person and a native speaker early and late in the dialog. Figure 1 shows the match values across the dialog for two female participants (BS and BR) with high scores for phonetic talent (respectively 1,4 and 1,5) and their partners.

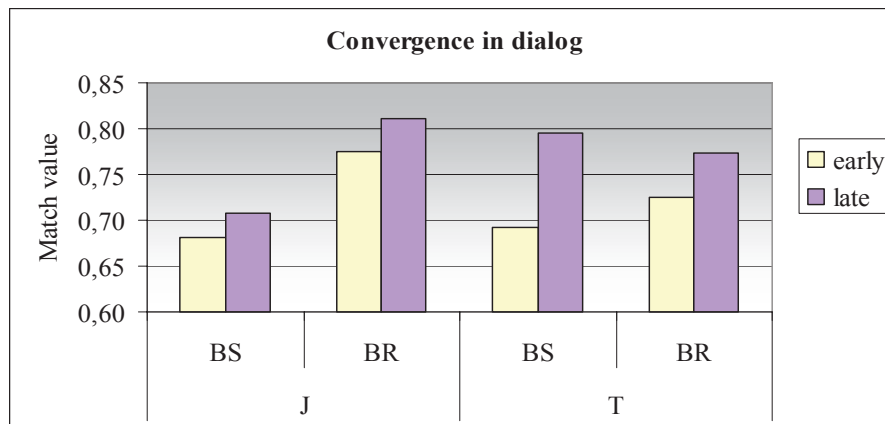


Figure 1. Convergence in dialog. Comparison of raw match values (mean across all target words) between an early and a late point in the conversations for two participants: BR and BS and two native speakers: J and T.

The match values are rising for all displayed conditions, indicating convergence in pronunciation between the participant and the respective English native dialog partner (J or T).

Table 1. Results of a bivariate correlation analysis between the word list task and the talent score.

		T2-T1	J3-J1
Talent	Pearson's correlation	0,893	0,880
	significance	0,107	0,120
	covariance	0,053	0,082

The second analysis (Table 1) aims at determining how persistent the convergent effect observed in the dialogs is. Compared were the mean match values between the word list readings before and after the dialogs, for four



subjects with varying talent scores (ranging from 4,2 up to 1,4) and the native speakers. The positive correlation results suggest that more talented subjects could indeed converge more than less talented speakers, although the effect did not reach statistical significance due to the small number of participants considered.

### Discussion

Although the described results are based on a limited amount of data and therefore of course have to be treated with caution, they nevertheless point to some promising tendencies that need to be further explored on the whole data set. Some important factors that were disregarded now (such as the direction of convergence in the dialog, possible gender differences or frequency effects within the target words) will then also be taken into account to provide a full picture of the relation between talent and convergence in native-nonnative interactions.

### References

- Bradlow, A. R., Baker, R. E., Choi, A., Kim, M. and van Engen, K. J. 2007. The Wildcat Corpus of Native- and Foreign-Accented English. *Journal of the Acoustical Society of America*, 121(5), Pt.2, 3072.
- Giles, H., Coupland, N. and Coupland, J. 1991. Accommodation Theory: Communication, context, and consequence. In H. Giles, J. Coupland and N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics*, 1-68. Cambridge, Cambridge University Press.
- Jilka, M., Anufryk, V., Baumotte, H., Lewandowski, N., Rota, G. and Reiterer, S. 2008. Assessing Individual Talent in Second Language Production and Perception. In: Rauber, A. S. et al. (Eds.). *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis, Federal University of Santa Catarina, 224-239.
- Johnson, K. 1997. Speech perception without speaker normalization: An Exemplar Model. In K. Johnson and J. W. Mullennix (Eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.
- Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382-2393.
- Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam, Benjamins.
- Wade, T., Dogil, G., Schütze, H., Walsh, M. and Möbius, B. (submitted). Syllable frequency effects in a context-sensitive segment production model.
- Wedel, A. and Van Volkinburg, H. (submitted): Modeling simultaneous convergence and divergence of linguistic features between differently-identifying groups in contact.

# **A comparison of Taiwanese sign language and manually coded Chinese: word length and short-term memory capacity**

Hsiu-Tan Liu<sup>1</sup>, Chin-Hsing Tseng<sup>2</sup> and Chun-Jung Liu<sup>1</sup>

<sup>1</sup>Chung Shan Medical University, Taiwan, R.O.C.

<sup>2</sup>National Kaohsiung Normal University, Taiwan, R.O.C.

## **Abstract**

Taiwanese Sign Language (TSL) is the natural language among deaf communities in Taiwan. Manually Coded Chinese (MCC) is the official instructional language. Previous studies have shown that the deaf students have great difficulty in comprehending stories in MCC, plausibly due to greater word length in MCC, which in turn may impair recall of MCC words. In Study I, deaf and hearing signers produced signs for 100 words in both MCC and TSL, and the word length was calculated for each sign pairs. It was found that MCC words were greater in length than the TSL words, whether produced by a hearing or a deaf signer. In Study II, the short-term memory capacity in the deaf signers was compared between word lists in TSL and in MCC. The participants were 44 senior high students in the deaf school and 20 deaf adults. The results showed that for deaf students and adults, the short-term memory capacity was inferior for the MCC list than for the TSL list, confirming our hypothesis.

Key words: Taiwanese sign language, manually coded Chinese, word length, short-term memory

## **Introduction**

In Taiwan, sign language is divided into two categories: Taiwanese Sign Language (TSL) and Manually Coded Chinese (MCC). TSL is the natural sign language used among the deaf communities. MCC is the official instructional language. It lays stress on the expression based on sequence of spoken language. Its vocabulary is developed by sign language research and development team.

Deaf communities did not accept MCC quite well. The deaf students only used MCC in class, but tend to use TSL during their leisure time outside of class. The previous studies found that the deaf student could comprehend the stories in TSL, but had great difficulty in comprehending the ones in MCC.

Why could deaf students not comprehend the stories in MCC easily? Supalla (1991), Hoffmeister (2000), and many other scholars believed that Manually Coded English combined the sign of visual language and the grammar of audible language, which may cause comprehending difficulties and cognitive memory load for the deaf students. We believe that the

sequential expression of MCC based on spoken language and the way of making signs one by one would cost much time and lead to extremely long word length, which might cause load on the deaf people's cognitive memory. That is the reason why they could not comprehend the stories in MCC easily.

This research assumes that the word length of MCC is greater than that of TSL, and the short-term memory capacity is inferior to that of TSL. In this research, Study I is the practical comparison between the word length of MCC and that of TSL, while Study II is the comparison between the short-term memory capacities of both languages.

### Study I

The major purpose of this study is to compare the word length of TSL and MCC. The experimental design used two-way ANOVA. The independent variables were the hearing status, and the sign language categories. The dependent variable was the word length of TSL and MCC produced by deaf signers and hearing signers. The participants in this study are 10 deaf adults and 10 hearing signers. The researcher sampled 100 words from sign language dictionary at random and required the participants to sign the words with these two sign languages respectively.

### Results

The average length of the 100 words signed by the participants in this study had been sorted and listed in Figure 1.

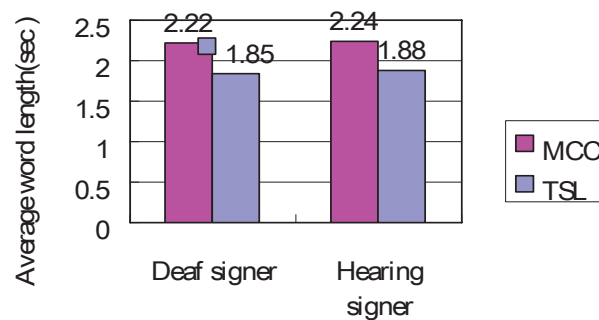


Figure1.Comparison of word length in MCC and TSL.

It was found from Figure 1 that MCC words length were greater than the TSL words both signed by hearing and deaf signers.

## Study II

The purpose of Study II is to compare the deaf signers' working memory capacities when they face the different morphologies. The independent variables of two-way ANOVA were age and morphology. The dependent variable was the sign language word memory capacity. The treatment levels of the age variable were deaf adults and deaf students. The morphology includes MCC borrowed words(MCC-B), MCC added pleonasms(MCC-A), MCC concatenated words(MCC-C), and TSL compounds(TSL). The MCC - C indicates that MCC borrows the words from TSL. It has only one sign, and it is signed identically by either MCC or TSL. In this study we called them TSL simple words (TSL-S). Both the MCC-A and MCC-C words were the sign language word combined with two signs. The TSL-C refers to the word with two signs in TSL. In the experimental material, the speeds of the four kinds of sign language words were adjusted to make no distinction exist in word length of the four kinds of sign languages. The participants were 44 senior high deaf students from Deaf School and 20 deaf adults.

## Results

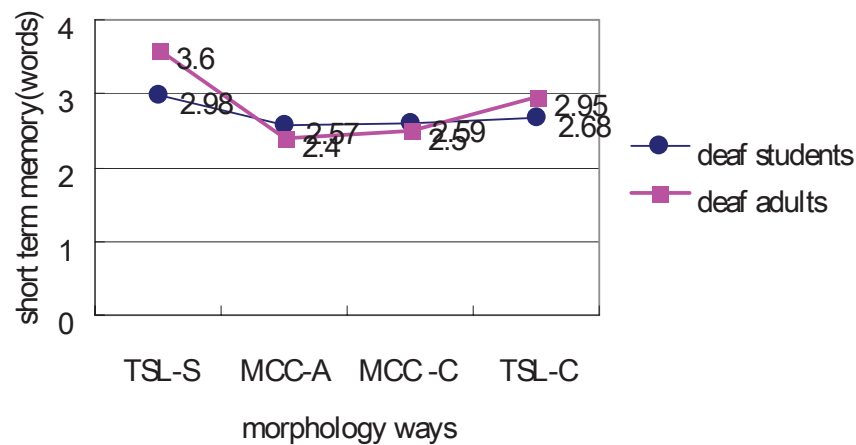


Figure2. Deaf signers' memory capacity of different morphology ways.

For the deaf students, TSL-S words memory extent was superior over MCC-A and MCC-C words.

For the deaf adults, the short-term memory capacity in TSL-S was superior than that in the other three kinds of sign language words (MCC-A, MCC-C, TSL-C), and the short-term memory capacity in TSL compounds was also superior than that in MCC-A and MCC-C.

### **Conclusion and discussion**

This research has proven that MCC is greater in word length than TSL, and the short-term memory capacity of deaf people in MCC words is inferior to that in TSL words.

The deaf people have a poor short-term memory capacity in MCC words, and that may have resulted from the effect of word length. However, Study II in this research still indicates that, for deaf students, the memory capacity was inferior for the MCC list than for the TSL list under the circumstance that the word length is under control. It indicates that other factors besides the word length may also have an effect of the deaf students' memory capacity when they memorize sign language words.

Memory is a key part in the cognitive process, so the educators should envisage the problem that deaf students have difficulties in memorizing MCC and resurvey the applicability of MCC to instruction.

### **References**

- Supalla, S. 1991. Manually coded English: The modality question in signed language development. In P. Siple, and S. Fischer (Eds.), *Theoretical issues in sign language research: Acquisition* (pp. 85-109). Chicago: University of Chicago Press.
- Hoffmeister, R. J. 2000. A piece of the puzzle: ASL and reading comprehension in deaf children. In C. Chamberlain, J. P. Morford, and R. I. Mayberry (Eds.), *Language acquisition by eye* (pp. 143-164). Mahwah, NJ: Lawrence Erlbaum Associates.

# The role of animacy in the production of Greek relative clauses

Sofia Loui and Silvia P. Gennari

Department of Psychology, University of York, U.K.

## Abstract

This paper examines the effects of animacy on relative clauses production in Greek. Previous studies in English suggest that animacy influences structure preferences in main and relative clauses: animate entities are typically made the subject of the verb and this tendency often results in passive structures. In the present experiment, Greek speakers were presented with pictures depicting actions with an animate and an inanimate participant or two animate participants and they were asked to write answers to questions about the inanimate or the animate patients of these actions. The results were then compared to the results obtained in English and Spanish. Results suggest that animacy does not play a critical role in determining Greek structures, as it does in English, indicating that language-specific constraints play a role in production mechanisms to a greater extent than conceptual factors such as animacy.

Key words: production, animacy, relative clauses, Greek

## Introduction

Research suggests that language production is an incremental process (De Smedt, 1990, Ferreira, 1996): speakers start uttering words before the whole message is ready to be produced and they typically start their utterances with information that is easier to retrieve from memory. Words and concepts that are more difficult to retrieve tend to be produced later in the sentence. It has been proposed that animate entities (typically humans) are conceptually more accessible than inanimate ones, as they are more easily retrieved from memory (Bock and Warren, 1985). Thus, animate entities tend to appear in earlier sentence positions such as subjects, and this tendency sometimes results in a preference for passive structures (Bock, 1987). For example, to express the message that a truck hit a boy, English speakers are much more likely to say *The boy was hit by the truck* than *The truck hit the boy*. This is because the tendency to locate animate entities in subject position determines that the sentence will be continued in a passive structure, even though such structures are relatively rare in English. Thus, in English, animacy affects syntactic structure via syntactic function assignment and is linked to the production of passives when certain messages are conveyed.

### Previous work on relative clauses production

Previous work also suggests that production of subordinate clauses such as relative clauses is also affected by animacy. Relative clauses provide information to identify a given participant for cases in which a noun phrase alone would not be sufficient. For example, in a situation in which there are two women, as in the present figure, the phrase *the woman* would not be sufficient to identify the woman being talked about, so a noun phrase modified by a relative clause such as *the woman who is punching the bag* would be required. For these structures, Gennari and Mac Donald (2008) argued that the head of the relative clauses is fixed in the initial sentence position due to discourse considerations. But the word order within the relative clause may vary depending on the function of the head nouns with respect to the relative clause verb, as in *The woman who is punching the bag* (*woman* = subject) vs. *the man that the woman is punching* (*man* = object). To evaluate the role of animacy in relative clause production, Gennari, Mircovic and MacDonald (2005) presented English participants with pictures showing actions with an inanimate theme (a woman punching a bag) or an animate patient (a woman punching a man), as in the figure above, and they were asked questions relevant to the theme or the patient of the action (*What is orange?* Or *Who is bold?*). Results showed that when the question was about the animate patient, 96% of the responses were passive structures (e.g., *The man who is being punched by the woman*) but when the question was about the inanimate theme, only 48% were passives. Thus, for inanimate head nouns, many actives structures were produced (e.g., *The bag that the woman is punching*). These results indicate that in English, the animacy of the entity being talked about influences the type of structure (active or passive) that is eventually produced in relative clauses: animate entities are more likely to be produced as subjects of the relative clause verb and thus result in a passive structure.



### Differences in production preferences across languages

One interesting question about these production preferences is whether languages with more flexible word order would also show preference for passive structures. In English, passives are necessary because subjects typically precede the verb, so when a noun has been uttered in initial position (due to its being more accessible), the only possible structural continuation is a passive structure, as in *the boy was hit by the truck* (*the boy*,

*the truck hit* is not acceptable in English). Indeed, it has been shown that in languages like Spanish with more flexible word order than English, the tendency to locate animate entities in subject position is modulated by language-specific constraints, e.g., speaker may use dislocated active structures such as *the boy, the truck hit* (OSV order instead of the SVO canonical order) (Prat-Sala and Branigan, 2000). In these structures the word order changes, without any change in the syntactic function of the entities. Nevertheless, Spanish, like English, shows a preference for passive structures in relative clauses when talking about animate entities, although this preference is smaller than in English (only 56% of animate responses were passives) (Gennari, Mircovic and MacDonald, 2005). Spanish also allows the use of impersonal constructions like *the man to whom (they) are punching* in relative clauses, making the continuation into a passive structure less likely. These findings suggest that conceptual accessibility constraints like those coming from animacy may manifest differently in different languages, depending on the grammatical possibilities of each language: when alternative structures are available, as in Spanish, the preference for passive is reduced compared to English.

The work presented here aims to explore the effects of animacy in the production of relative clauses in Greek, a language that shows flexibility in word order, such that any arrangement of subject, verb and object is grammatically licenced and independent of syntactic function. This study will indicate what structural possibilities are available in Greek and will help determine how language-specific rules interact with conceptual aspects of production processes such as those coming from animacy.

### **Greek production in relative clauses**

30 Greek native speakers aged 18-35 years participated in the present experiment conducted via the web. They were presented with 20 pictures where an animate and an inanimate or two animate entities take part in an action. There were two conditions. For each picture, each participant answered one question either about the animate patient (animate condition) or about the inanimate theme (inanimate condition) (see example above). All participants were asked equal number of questions from both conditions and they produced their answers in written form. Participants were advised to focus on the action of the sentence and not on the theme's or patient's location in the picture, their clothes or colours. 40 filler items were also included in the experiment that encouraged participants to produce different types of clauses, other than relative ones. Before the experimental trials, participants were provided with examples. The responses were coded by active versus passive relative clause structures. The pictures and procedure of this study were the same as in the previous English and Spanish studies.



The results indicate that there is an overall preference in Greek speakers to use more active than passive structures in both conditions. They used 87% and 97% active structures for the animate and the inanimate condition respectively. Hence, they were more likely to produce active structures like *The man/the bag that the woman is punching* than passive structures like *The man/the bag that is being punched by the woman*. However, they were more likely to produce passive structures when answering the questions about animate patients (animate condition) than about inanimate themes (inanimate condition). They showed a 13% significant preference in producing more passive structures in the animate condition compared to 3% of passive structures production in the inanimate condition. This preference is much smaller than in Spanish (56%) and even smaller than in English (96%).

These results suggest that language-specific constraints play a role in language production. The effect of animacy on sentence structure, which is associated with conceptual accessibility and ease of retrieval, is minimal in Greek, whereas it is critical in English structural preferences. It is possible that in Greek, passive structures are not really accessible to speakers so that active structures must be produced. This makes language-specific constraints more critical than conceptual factors in determining sentence structure. However, if animacy matters, this may come at a greater cost in planning the utterance and thus may require more planning time. This possibility is currently being investigated.

## References

- Bock, J.K. 1987. Coordinating words and syntax in speech plans. In A. Ellis (ed) *Progress in the pathology of language*. London: Erlbaum.
- Bock, J.K. and Warren, R.K. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47-67.
- De Smedt, K.J. 1990. IPF: An incremental parallel formulator. In R. Dale, C. Mellish and M. Zock (Eds.), *Current research in natural language generation*, Cognitive Science Series (pp. 167-192). San Diego: Academic Press.
- Ferreira, V.S. 1996. Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35, 724-755.
- Gennari, S.P., McDonald, M. C. 2008. Linking production and comprehension processes: The case of Relative clauses. submitted.
- Gennari, S.P., Mirkovic, J. And MacDonald, M. C. 2005. The role of animacy in relative clause production. Paper presented at the 17th Annual Cuny Conference on Human Sentence Processing.
- McDonald, J.L., Bock, J.K. and Kelly, M.H. 1993. Word and world order: Semantics, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25, 188-230.
- Prat-Sala, M. and Branigan, H.P. 2000. Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, 42(2), 168-182.

# Acoustic model of stress in standard Greek and Greek dialects

Anastassia Loukina

Phonetics Laboratory, University of Oxford, UK

## Abstract

The paper compares acoustic models which predict the position of stress in bisyllabic words in Greek varieties. It is shown that while in Thessalian Greek ratios of peak amplitude and F1 are the best predictors of stress, in Athenian and Cypriot Greek these are ratios of durations and  $f_0$ . Furthermore in Cypriot Greek morphological category of the token affects the acoustic prominence of the stressed vowel.

Key words: phonetics, vowels, stress, Modern Greek, Greek dialects

## Introduction and Background

Traditionally languages with lexical stress have been divided into languages where prominence is cued primarily by fundamental frequency and languages where prominence was mainly linked to loudness. This division was later re-evaluated as a distinction between languages where pitch is the only acoustic correlate of stress (“non-stress”) vs. languages where as well as fundamental frequency other acoustic properties may also be used to achieve prominence (“stress”) (Beckman, 1986). However, even within the category of “stress” languages, there exists great variation in the acoustic means used to mark the prominence of a stressed vowel. The hierarchies of acoustic correlates of stress may differ not only between different languages, but also across several regional varieties of the same language. Thus variation in the acoustic correlates of stress has recently been reported for British English (Kochanski et al., 2005) and Dutch (Fournier et al., 2006).

The accentuation systems of Standard Modern Greek and Modern Greek dialects are very similar and are traditionally described as “stress accent”. Nevertheless, there is certain evidence that the acoustic correlates of stress in Greek dialects may differ from Standard Greek. Recent experimental studies have shown that stressed and unstressed vowels in Standard Modern Greek show consistent differences not only in amplitude, but also in duration. (Botinis, 1989, Arvaniti, 2000, Fourakis et al., 1999). Contrary to these results, Pernot (1907) found that in Chios Greek rising tone was more consistent correlate of stress than duration or amplitude. For Northern Greek dialects, Chatzidakis (1892) observed that stressed syllables become longer and the difference in duration is greater than in the Southern dialects. Northern Greek dialects also show reduction of unstressed mid vowels (cf.

Newton, 1972). Although in Athenian and Cypriot Greek, the distribution of vowels is not dependent on stress, acoustic studies showed that unstressed vowel in Standard Greek tend to have lower F1 than stressed vowels (Nicolaidis, 2003, Fourakis et al., 1999, Loukina, in press).

In this study I will attempt to explore the contribution of each of these parameters to the prominence of stressed vowels in Athenian, Thessalian and Cypriot Greek. The study is based on multiple occurrences of disyllabic words with the same vowel phonemes in both syllables (341 tokens). The tokens were extracted from spontaneous monologues recorded from speakers of the corresponding dialects in Athens, Thessaly and Cyprus and analyzed with speech-processing software.

## Results

In *Athenian Greek* words with stress on the first syllable, stressed vowels were often distinguished by a higher peak  $f_0$ , higher amplitude and also longer duration than the unstressed vowel, but sometimes by only two or one of these parameters. In words with stress on the second syllable in most cases, acoustic prominence of the stressed vowels was achieved by all three of these parameters, that is a difference in duration usually co-occurred with a difference in  $f_0$  and with a difference in amplitude. Stressed and unstressed vowels of the same phonemic category in some cases also differed in F1 although this depended on the vowel and phonetic context.

Binary logistic regression analyses showed that in Athenian Greek ratios of durations,  $f_0$  and amplitude between first and second vowel each allowed the position of the stress to be predicted correctly in about 80% of cases. The effect of F1 was also significant, although the prediction rate was smaller than for other measures (64%). For models based on combinations of different measures, the forward stepwise analysis showed that a combination of ratios of amplitude integral and  $f_0$  achieves better prediction rate (89.7%) than a combination of duration and  $f_0$  or amplitude and  $f_0$ . This agrees with results by Arvaniti (2000), who suggested that amplitude integral is a better indicator of stress in Greek than duration or peak amplitude. Further contribution of F1 was found to be insignificant.

In *Thessalian Greek* words with stress on the first syllable, the stressed vowel was in most cases distinguished by a higher amplitude than the unstressed vowel, usually accompanied by higher peak  $f_0$  and sometimes longer duration. In words with stress on the second syllable stressed vowels were most frequently distinguished by greater amplitude and longer duration, often accompanied by higher peak  $f_0$ . Stressed and unstressed /e/, /o/ and /a/ also showed consistent difference in F1.

Models based on ratios of peak amplitude, F1 and duration separately correctly predict the number of correct syllables at above the chance level.

The model based only on ratios of  $f_0$  performed at the chance level. The forward stepwise method showed that only the ratio of peak amplitudes and the ratio of F1 are good predictors (91.3%), while ratio of durations and ratio of  $f_0$  do not improve the model. This provides experimental support for the numerous impressionistic observations that stress in Northern dialects may be “stronger” or “more dynamic” than in Athenian Greek (Chatzidakis, 1892, Tzartanos, 1909). Unlike Athenian Greek, in Thessalian Greek substitution of the ratio of peak amplitudes by the ratio of amplitude integrals did not lead to any major changes in the model and the effect of peak amplitude still remained significant.

In *Cypriot Greek* words with stress on the first syllable, the stressed vowel usually had a higher amplitude and a higher peak  $f_0$ , but it was often shorter than the unstressed vowel. In words with stress on the second syllable the stressed vowel was longer than the unstressed vowel but there was no difference between amplitudes or peak  $f_0$  of the vowels. There was no consistent difference in F1 between stressed and unstressed vowels.

Models based on ratios of amplitude, duration or  $f_0$  separately correctly predict the position of stress in about 65% of cases. The model based on ratios of F1 failed to predict correct position of stress above the chance level. For models based on a combination of parameters, the model based on  $f_0$  and amplitude integral achieved a similar result to the model based on amplitude, duration and  $f_0$  and correctly classified 73.6% of cases.

The relatively poor performance of the classifier is due to the fact that words in Cypriot Greek in general showed similar acoustic patterns regardless of stress location: peak amplitude and peak  $f_0$  on the first vowel and maximum duration on the second vowel. Changes to this pattern would usually indicate that this vowel is stressed.

There were also a number of cases where the stressed vowels probably were not assigned acoustic prominence or acoustic prominence was expressed by other means not considered in this analysis. Analysis of Cypriot Greek data showed stressed vowels in nouns were more likely to be acoustically prominent than the stressed vowels in adjectives or adverbs. As a result, the predictive power of the model in Cypriot Greek could be substantially increased by introducing morphological category (“noun vs. other”) as one of the parameters. This model correctly predicted the location of stress in 79.3% of cases. The introduction of morphological category did not improve the model in the other two varieties.

Further observations refer to the overall success rate of the models. In Thessalian Greek even less successful models achieved prediction rate of 80% if they included peak amplitude or amplitude integral. This suggests that amplitude is a very robust acoustic correlate of stress in this variety. In Cypriot Greek even the best performing model correctly predicted only 80%

of the data. These results corroborate previous impressionistic observations by Menardos (1894) that Cypriot Greek allows several words to be combined into one phrase with one main stress.

### Conclusion

Analysis of Modern Greek dialects also provides further evidence for heterogeneity within the traditional category of “stress” languages. The comparison between the acoustic correlates of stress in three varieties of Modern Greek confirmed that lexical stress is not always associated with acoustic prominence and the acoustic stability of stress may differ across the varieties. These varieties present very few differences in the position of stress, which suggests that the acoustic correlates of stress are not necessarily related to the phonological function of stress.

### References

- Arvaniti, A. 2000. The phonetics of stress in Greek. *Journal of Greek Linguistics* 1, 9-39.
- Beckman, M. 1986. *Stress and non-stress accent*, Dordrecht, Foris publications.
- Botinis, A. 1989. *Stress and prosodic structure in Greek*: Lund University Press.
- Chatzidakis, G. N. (1892) *Einleitung in die neugriechische Grammatik*, Leipzig, Breitkopf und Härtel.
- Fourakis, M., Botinis, A. and Katsaiti, M. 1999. Acoustic characteristics of Greek vowels. *Phonetica* 56, 28-43.
- Fournier, R., et al. 2006. Perceiving Word Prosodic Contrasts as a Function of Sentence Prosody in Two Dutch Limburgian Dialects. *Journ. of Phonetics* 34, 29-48.
- Kochanski, G., et al. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118, 1038-1054.
- Loukina, A. (in press) Phonetic variation in regional varieties of Modern Greek: vowel raising. *Proceedings of the 3rd International Conference on Modern Greek Dialects and Linguistic Theory*. Nicosia, 14-16 June 2007.
- Menardos, S. 1894. Φωνητική του διαλέκτου των σημερινών Κυπρίων. *Αθήνα*, 6, 146-173.
- Newton, B. 1972. *The generative interpretation of dialect*. Cambridge University press.
- Nicolaidis, K. 2003. Acoustic variability of vowels in Greek spontaneous speech. 15th ICPHS. Barcelona.
- Pernot, H. O. 1907. *Études de linguistique néo-hellénique*, Paris.
- Tzartzanos, A.A. 1909. *Περί της συγχρόνου Θεσσαλικής διαλέκτου*, Αθήναι, Τυπ. Π.Α. Πετράκου.

# Using F2 transition parameters in distinguishing Persian affricates from homorganic consonants

Zahra Mahmoodzade and Mahmoud Bijankhan  
Linguistics Department, University of Tehran, Iran

## Abstract

F2 transition parameters including F2 onset and offset frequency and locus equations have been studied for Persian affricates /dʒ, tʃ/ and homorganic dental plosives /d, t/ and post alveolar fricatives /ʒ, ʃ/. The F2 offset frequencies hierarchy is follow: post alveolar fricatives > affricates > dental plosives. This parameter makes a significant difference among mentioned consonants; however it does not show voiced-unvoiced contrast. F2 onset does not provide regular results and locus equations coefficients do not show a significant difference between affricates and homorganic consonants.

Keywords: Persian, affricates, homorganic consonants, F2 transition parameters

## Introduction

Vocalic parameters in consonant-vowel boundaries are important acoustic cues in consonant identification. Dorman et al. (1980) found that vocalic portion is an important acoustic cue in fricative-affricate contrast in word final position. In the previous studies, we investigated different acoustic parameters for making contrast between Persian affricates /dʒ, tʃ/ and homorganic dentals /d, t/ and fricatives /ʒ, ʃ/ (Heselwod et al. 2008, Mahmoodzade and Bijankhan 2007). In this paper, F2 transition parameters including F2 onset and offset frequencies and also locus equations in the vocalic portion have been selected for studying the mentioned contrasts. These parameters are investigated mainly as acoustic cues demonstrating place of articulation differences in stops (Sussman et al. 1991, Kewley-Port 1982) and also fricatives (Jongman 2000, Whalen 1991). Some experiments used them for differences in manner of articulation too (Fowler 1994, Dorman et al 1980).

## Elicited Data and Measurements

Ten male native speakers of Standard Persian (Farsi), aged between 20-40 years, were recorded producing isolated word forms with affricates /dʒ, tʃ/ and homorganic fricatives /ʒ, ʃ/ and plosives /d, t/ in the vocalic contexts /'-aC₂/, /Ca'-aC/ and /'Ca-/ in initial, medial and final positions. Four repetitions of the word list were produced. There were 720 tokens altogether. The data was sampled at the rate of 22050 Hz. The recording and analyzing

of data was done by using CSL 4400 Main Program. Both post and pre-consonantal F2 transitions parameters were studied. ANOVA test (SPSS package 13) used for statistical analyses.

F2 onset Frequency is the frequency of second formant at the starting point of transition and was measured for affricates and homorganic consonants at the initial and medial position by calculating formants history on the spectrogram and determined at point after the release phase of consonants. The release of affricates /dʒ, tʃ/ and plosives /d, t/ is accompanied with friction and aspiration (Heselerwood et al. 2008), therefore the onset was determined at end of friction before aspiration. In the case of voiced unaspirated stops, the onset was determined at the first glottal pulse after the release. For fricatives /ʒ, ʃ/, the onsets were determined at the first glottal pulse after frication phase following Sussman and Shore (1996). FFT spectrum was calculated and F2 onset was measured manually.

F2 offset frequency is the frequency of second formant at the end of transition. The method for measuring F2 offset frequency is like F2 onset frequency, except the locus was determined for affricates and plosives at the last glottal pulse before closure and for fricatives before frication in medial and final position.

Locus equations “cue place indirectly by quantifying directly the degree of coarticulatory overlap (coarticulation resistance) between consonant and vowel” (Fowler, 1994). Locus equations coefficients, slope and y-intercept were measured following sussman et al. (1991) by plotting F2 onset or offset frequencies on the y-axis and F2 steady state frequencies on the x-axis. F2 steady state was determined visually following Sussman et al. (1991) and measured like F2 onset or offset frequencies. The regression lines traced on these points and the correlation coefficient (R) and regression functions coefficients calculated by using SPSS software.

## Results and Discussion

The mean values of F2 offset frequencies are given in table 1. ANOVA tests show that this parameter can make a significant difference between consonants in medial ( $F_{(5, 234)} = 26.4, p < 0.05$ ) and final position ( $F_{(5, 234)} = 40.7, p < 0.05$ ). LSD Post Hoc tests show the following same results both for medial and final positions: 1) There is not a significant difference between /dʒ- tʃ/, /d- t/, /ʒ- ʃ/ pairs ( $p > 0.05$ ). 2) There is a significant difference between /dʒ- d/ and /dʒ- ʒ/ pairs ( $p < 0.05$ ). 3) There is a significant difference between /tʃ- t/ and /tʃ- ʃ/ pairs ( $p < 0.05$ ). 4) There is a significant difference between /ʒ- d/ and /ʃ- t/ pairs ( $p < 0.05$ ).



Table 1. Offset frequencies results in Hz.

Consonants	Medial Position		Final Position	
	Mean	St. Deviation	Mean	St. Deviation
dʒ	1749	87.4	1733	95.9
tʃ	1760	89.5	1740	90.1
d	1598	101.9	1586	86.6
t	1552	186.7	1587	74.7
ʒ	1822	221.3	1848	149.2
ʃ	1841	134.8	1819	144.6

The statistical analysis of F2 onset frequencies in initial and medial position does not provide regular results for relying on. Locus equations coefficients do not show a significant difference between the mentioned consonants in initial, medial and final positions, both in post and pre-consonantal positions.

Results demonstrated that only pre-consonantal F2 offset frequencies can make a significant difference between Persian affricates and homorganic consonants. This parameter shows both contrasts in place and manner of articulation. There is a significant difference between F2 offset frequencies before closure and friction. One reason that F2 onset frequencies do not provide such a classification could be that F2 onset was measured after release burst therefore frequency differences decrease at this time as a result of closure opening. The mean F2 offset frequencies hierarchy is as follow: post alveolar fricatives > post alveolar affricates > dental plosives. Wilde (1993) and Jongman et al. (2000) found that in a given vowel context, F2 onset is progressively higher as the place of constriction moves back in the oral cavity. Our results with offset frequencies show that F2 offset frequencies of /ʒ, ʃ/ is higher than /dʒ, tʃ/, therefore the place of articulation of fricatives is further back in relation to affricates. Another matter is that there is no significant difference in F2 offset frequencies between voiced and voiceless pairs; this is the same result that Jongman et al. (2000) found by comparing F2 onset frequencies between voiced and voiceless labiodental, dental and alveolar English fricatives.

## References

- Dorman et al. 1980. Acoustic cues for a fricative-affricate contrast in word-final position. *J. Phonetics*, Vol. 8, 397-405.
- Fowler, C. 1994. Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Percept. Psychophys.* 55(6), 597-610.
- Heselwood, B. and Mahmoodzade, Z. 2007. Vowel onset characteristics as a function of voice and manner contrasts in Persian coronal stops. *Leeds working Papers in linguistics and Phonetics*, Vol. 12, 125-142.



- Heselwood, B., Mahmoodzade, Z., Bijankhan, M. 2008. Phonetic differences distinguishing Persian coronal plosives from affricates between release and vowel onset. BAAP, Sheffield, 31 March - 2 April.
- Jongman et al. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108(3), 1252-1263.
- Kewley-Port, D. 1982. Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *J. Acoust. Soc. Am.* 72(2), 379-389.
- Mahmoodzade, Z and Bijankhan, M, 2007. Acoustic analysis of the Persian fricative-affricate contrast. ICPHS XIV, Saarbrücken, 6-10 August.
- Sussman, H. M., McCaffrey, H. A., Matthews, S., A. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *J. Acoust. Soc. Am.* 90(3), 1309-1325.
- Sussman, H. M. and shore, J. 1996. Locus equations as phonetic descriptors of consonantal place of articulation. *Percept. Psychophys.* 58, 936-946.
- Whalen, D. H. 1991. Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *J. Acoust. Soc. Am.* 90(4), 1776-1785.
- Wilde, L. 1993. Inferring articulatory movements from acoustic properties at fricative vowel boundaries. *J. Acoust. Soc. Am.* 94, 1881.

# Intonation of parentheses in spontaneous French sentences

Philippe Martin

EA333 ARP, UFRL, Université Paris Diderot, France

## Abstract

Traditionally, parentheses have been described by phoneticians as speech segments with lower intensity, restrained pitch variation and speech rate faster than the rest of the sentence. This iconic coding may be found in some examples of read (prepared) speech, but spontaneous speech data tend to show that specific prosodic mechanisms do exist to either 1) integrate the syntactic parenthesis into the overall prosodic structure of the sentence or 2) isolate it with an independent prosodic structure. The realizations chosen by speakers appear to be dictated by the interaction of semantic, syntactic and prosodic markers to ensure the parentheses will be identified as such by listeners.

Key words: intonation, parenthesis, prosodic structure, macrosyntax

## Theoretical background

In the theoretical background developed by GARS in Aix-en-Provence, a sentence, in spontaneous speech, can be analyzed in 5 distinct macrosyntactic types: 1) the Kernel (Noyau), which can form a syntactically and prosodically well formed autonomous speech unit, and optionally 2) one or more Prefixes, preceding the Kernel; 3) one or more Parentheses, imbedded in a Prefix or in the Kernel; 4) one or more Postfixes and 5) one or more Suffixes. Postfixes and Suffixes are both placed after the Kernel, and differ by the nature of the dependency relation which links them to the Kernel: Postfixes use a prosodic dependency relation manifested by specific contours on their final stressed syllables (flat contours in the declarative case), whereas Suffixes are linked to the Kernel with a syntactic, or possibly semantic, relation. In this arrangement, a canonical parenthesis is a verbal construction which interrupts the progression of another verbal construction or a sequence of verbal constructions (Blanche-Benveniste, 1997).

Various types of syntactic parentheses can be considered:

- a) Inside a macrosegment stress group: in the sequence *est-ce que nous sommes **je lance le débat du même coup hein** est-ce que nous sommes euh prêts?* [Retraites] the parenthesis **je lance le débat du même coup hein** is inserted in the stress group *nous sommes prêts* between the auxiliary *sommes* and the adjective *prêts* and is followed by a copy of the segment interrupted *est-ce que nous sommes*;

- b) Between stress groups of a macrosegment (Kernel, Prefix, Postfix or Suffix), as in the example *la fonction publique [...] c'est pas du tout l'emploi à vie **la mobilité professionnelle est très forte dans la fonction publique** mais c'est le salaire à vie [retraites]*, where the Parenthesis is located inside the Kernel *c'est pas du tout l'emploi à vie* mais *c'est le salaire à vie*, but not inside a Kernel stress group;
- c) Between two macrosegments. In the latter configuration it is often difficult to differentiate the parenthesis from a Prefix if it appears before the Kernel and from a Suffix when it occurs after the Kernel. In the example, without intonation, the parenthesis can also be considered as a Prefix: *après tu ouvres les feuillets comme un livre et avec la lumière du soleil **il faut absolument qu'il y ait le soleil** + il y a une ombre portée de la colonne vertébrale qui te dit que là il y a un poisson ou autre chose [fossiles]*.

Macrosegments are syntactically “floating” in the sentence, i.e. no syntactic dependency relation organizes their possible hierarchy, which is defined by intonation, i.e. by the prosodic structure. Furthermore, the Kernel is always associated to a well formed prosodic structure. This is due to the nature of prosodic dependency relations in French where prosodic markers signal a dependency to the right. Therefore the Kernel constitutes a well formed sequence both syntactically and prosodically. This can be tested experimentally by isolating a Kernel macrosegment with a sound editor, as the resulting sequence should be perceived by itself as complete and well formed independently.

As seen above, Parentheses are macrosegments inserted somewhere either inside a stress group, between stress groups in a macrosegment, or between macrosegments. Aligned with text parentheses, prosodic segments exist either independent or integrated in the overall prosodic structure of the sentence. In the last case, these segments end with a rising continuation contour, whereas independent segments end with a falling terminal contour.

### Prosodic parentheses

Debaisieux and Martin (2007) conducted an experimental analysis based on the macrosyntactic approach described above, hunting for totally well formed and complete prosodic structures which could be inserted inside the Kernel. Despite the relatively large number of experimental data analyzing the final prosodic movement in some 200 spontaneous speech examples, only 22 examples showed a final falling melodic contour, the majority of the other sentences exhibiting a rising contour. No clear correlation could be found between rising or falling contours ending the parentheses and the nature of the conjunction that introduces them (*si, parce que, comme,*

*puisque, enfin*), or the lack of such introductory unit. No correlation was found either with the position of the parenthesis inside the sentence (inside stress groups, between stress groups, between macrosegments).

### Combinatorial analysis

If we consider the sentence organized into two syntactic and prosodic lines, some syntactic, semantic or prosodic markers must operate at both ends of the parenthetical macrosegments to indicate its nature to the listeners. These markers could appear either on the preceding macrosegment or at the beginning of the parenthesis to indicate the left boundary, and at the end of the parenthesis or at the beginning of the following macrosegment to mark its right boundary (Fig. 1).

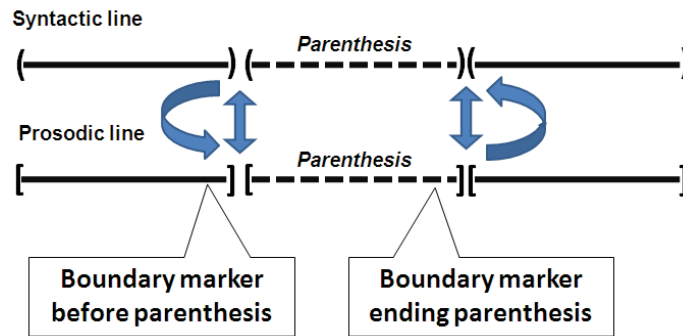


Figure 1. Syntactic and prosodic markers of parenthetical macrosegments.

Representing by A, B and C the text line units (unit B being the parenthesis), and a, b and c the prosodic units (unit b being the prosodic unit align with the text parenthesis), and denoting by X and x the presence of a marker located at the end of each text or prosodic unit, 8 cases are possible:

No prosodic marker:

A X B X C  
a b c

Complete or partial prosodic redundancy:

A X B X C      A X B X C      A X B X C  
a x b x c      a x b c      a b x c

Syntactic marker starting unit C (reprise):

A B X C      A B X C      A B X C  
a x b x c      a x b c      a b x c

No syntactic marker (parenthesis non introduced by a conjunction):

A B C  
a x b x c

Considering these combinatorial possibilities, it becomes clear that we cannot expect all syntactic and prosodic markers to be effectively present in a given realization. In particular, prosodic markers, mostly through the presence of a rising or falling melodic contour located on the last stressed syllable of the parenthetical macrosegment, simply reveal the integration of the parenthesis into the overall prosodic structure of the sentence, or on the contrary by a falling contour the presence of an independent prosodic structure imbedded in the overall prosodic structure of the sentence. If this approach is correct, prosodic markers will be always present if no other semantic or syntactic marker is present. Indeed data showed that 18 cases out of 22 parentheses ending with a falling contour are correlated with non introduced parentheses with no text reprisal after their end.

### **Conclusion**

Since no clear correlation has been found between the characteristics of parenthesis and their syntactic properties, in particular between their position in the macrosyntactic structure (inside the Kernel, between Prefixes, etc.) and the particular conjunction (or their absence) that introduced them, another view must be considered. As prosodic parentheses can function independently from the syntactic parentheses, our data suggest that the congruence between the prosodic and syntactic parentheses is only obligatory when no other markers, syntactic or semantic, are used by the speaker.

### **References**

- Blanche-Benveniste, C. 1997. *Approches de la langue parlée en français*. Paris : Ophrys.
- Debaisieux, J-M. et Deulofeu, H-J. 2006. Cohérence et syntaxe : le rôle des connecteurs. in Frédéric Calas, (éd) *Cohérence et discours*. PUPS. Paris, 197-207.
- Debaisieux, J-M. et Martin, Ph. 2008. Les parenthèses : étude macrosyntaxique et prosodique sur corpus. M. J. Beguelin and M. Avanzi (ed.) *Actes du I<sup>o</sup> Colloque international de macro-syntaxe "La parataxe"* (Neuchâtel 12-15 février 2007), Tübingen: Niemeyer.
- Gachet, F. et Avanzi, M. 2008. Les parenthèses en français parlé : étude prosodique. Communication aux Journées Conscila, « Les parenthèses en français », 14 mars 2008, Paris, ENS Ulm.
- Martin, Ph. 2006. Intonation du français: parole spontanée et parole lue, *Estudios de Fonética Experimental*, Vol XV, 2006, Barcelona, 133-162.

# Perception of consonant clusters in Japanese native speakers: influence of foreign language learning

Hinako Masuda and Takayuki Arai

Department of Science and Technology, Sophia University, Japan

## Abstract

Previous research on the perception of consonant clusters by Japanese native speakers has revealed that they are highly likely to perceive a vowel between the two consonants even if there are no vowels inserted (Dupoux et al. 1999). The present study further investigates this issue but by dividing the group of Japanese native speakers into two groups: Japanese-English bilinguals and Japanese monolinguals. This study aims to look into the influence of language learning experience on the perception of foreign sounds. Results of the perception test revealed that bilinguals made less error than monolinguals in identifying pseudowords with and without consonant cluster.

Key words: consonant cluster, epenthetic vowel, bilingual, speech perception

## Introduction

Numerous previous studies have shown that one's native language has a great effect on the perception of foreign sounds (Miyawaki et al., 1975; Peperkamp et al., 1999; Takagi, 2002; Sebastian-Galles, 2005). Previous study by Dupoux et al. (1999) has revealed that Japanese native speakers are more likely to perceive an 'illusory vowel' between two consonants even if no vowels are inserted. This is due to the difference in the phonotactic structure of French and Japanese. While French allow consonant clusters, Japanese does not; therefore, Japanese apply the rules of the Japanese phonotactics when hearing foreign sounds.

Consonant clusters are also common in English. Below are some English words introduced to the Japanese language. Japanese often insert the vowel [u] between consonants to avoid consonant clusters (Nishimitsu, 2004).

kiss [kis] : /kisu/	milk [milk] : /miruku/
cat [kæt] : /kjat.to/	ice cream [aiskri:m] : /aisukuriimu/

In the present study, we aim to investigate the perception of pseudowords with and without consonant clusters by dividing the Japanese native speaker group into two: Japanese-English bilinguals and Japanese monolinguals. By investigating the difference of perceptual ability, we will be able to see how different language learning backgrounds affect foreign language perception.

## Experiment

The present experiment aims to investigate the difference in the perceptual ability between two groups of native speakers of Japanese: Japanese-English bilinguals and Japanese monolinguals.

## Participants

17 Japanese-English bilinguals, 12 women and 5 men, and 22 Japanese monolinguals, 8 women and 14 men, participated in this experiment.

All bilinguals have experience of living in an English-speaking country and receiving education in an English-speaking school for at least 2 years. None of the monolinguals, on the other hand, have experience of overseas for more than one month.

The ages ranged from 19 to 25 for bilinguals (average of 23.4 years), and 18 to 25 for monolinguals (average of 20.8 years). The length of overseas experience for bilinguals ranged from 2 to 8.5 years (average of 5.7 years). None of the participants reported any hearing problems.

## Stimuli

The speaker of the recorded stimuli used in the perception experiment is a Japanese-French bilingual speaker. Her first language is Japanese, and is highly fluent in both languages. She is also an experienced phonetician.

36 pseudowords (18 pairs of pseudoword sets with and without consonant clusters) were produced in the sentence “Dites (pseudoword) deux fois.” [Say (pseudoword) two times.] Each word was recorded three times. 16 pairs of pseudowords were used as stimuli for the experiment, and the remaining 2 pairs (abmo/abumo, ebzo/ebuzo) were used for practice trials. The pseudowords are listed in Table 1.

Table 1. The list of 18 pairs of pseudowords (36 pseudowords) with and without consonant clusters.

VCCV	VCVCV	VCCV	VCVCV
abge	ab <u>u</u> ge	eshmo	esh <u>u</u> mo
abmo	ab <u>u</u> mo	ibdo	ib <u>u</u> do
agmi	ag <u>u</u> mi	igna	ig <u>u</u> na
akmo	ak <u>u</u> mo	ikma	ik <u>u</u> ma
ashmi	ash <u>u</u> mi	ishto	ish <u>u</u> to
ebza	eb <u>u</u> za	obni	ob <u>u</u> ni
ebzo	eb <u>u</u> zo	ogza	og <u>u</u> za
egdo	eg <u>u</u> do	okna	ok <u>u</u> na
ekshi	ek <u>u</u> shi	oshta	osh <u>u</u> ta

## Procedure

The participants were instructed to listen to the 3 words through headphones, and to judge whether the second word was more similar to the first or the third word. The order of the presented words (VCCV-VCVCV order or VCVCV-VCCV order) was randomized. This identification experiment consisted of 192 trials (32 words x 2 orders x 3 sets of recorded words) after 8 practice trials. The participants received no feedback on the practice trials.

## Results

The Mann-Whitney test found a significant difference between the two groups of participants ( $p < 0.05$ ). Table 2 shows the average, range, and percentages of the number of errors made by each participant group.

Table 2. The average and range of the number of errors (% of errors) made by bilinguals and monolinguals. The Mann-Whitney test found a significant difference between the two groups ( $p < 0.05$ ).

	Bilingual (n=17)	Monolingual (n=22)
Average	3.8 (1.9%)	13.1 (6.8%)
Range	0-12 (0-6.2%)	2-35 (1-18.2%)

## Discussion

The perception experiment conducted in the present research aimed to find the difference in the identification ability between Japanese-English bilinguals and Japanese monolinguals. The stimuli used were pseudowords with and without consonant clusters. The results of the experiment showed that there was a significant difference in the number of errors made by the two participant groups.

The result of the experiment revealed that there is a difference in the perceptual ability of consonant clusters depending on whether the participants have experience of living abroad. However, the percentages of correct answers were above 90% for both groups. This may have been caused by the stimuli used. The present research used stimuli that were neither digitally processed nor spoken fast in speed. The difference between the two groups' results may be larger if the stimuli had been produced in a more natural, faster speed instead of slow speed which may unconsciously cause vowel epenthesis. Further research will be performed with improved stimuli.



### Acknowledgments

This study work was partially supported by Sophia University Open Research Center from MEXT.

### References

- Dupoux, E., Kakehi, K., Hirose, Y., and Pallier, C. 1999. Epenthetic Vowels in Japanese: a Perceptual Illusion? *Journal of Experimental Psychology: Human Perception and Performance*, Volume 25, Number 6, 1568-1578.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., and Fujimura, O. 1975. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics* 18, 331-340.
- Nishimitsu, Y. (ed) 2004. *Nishieigo taisho ni yoru eigogakugairon zouban*. Tokyo, Kuroshio Publishers.
- Peperkamp, S., Dupoux, E., and Sebastian-Galles, N. 1999. Perception of stress by French, Spanish, and bilingual speakers. *Eurospeech '99 Proceedings, ESCA 7th European Conference on Speech Communication and Technology*, vol.6, 2683-2686.
- Sebastian-Galles, N. 2005. Cross-Language Speech Perception. *The Handbook of Speech Perception*, 546-566, Australia, Blackwell Publishing.
- Takagi, N. 2002. The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America* 111(6), 2887-2896.

# How are words reduced in spontaneous speech?

Holger Mitterer

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

## Abstract

Words are reduced in spontaneous speech. If reductions are constrained by functional (i.e., perception and production) constraints, they should not be arbitrary. This hypothesis was tested by examining the pronunciations of high- to mid-frequency words in a Dutch and a German spontaneous speech corpus. In logistic-regression models the "reduction likelihood" of a phoneme was predicted by fixed-effect predictors such as position within the word, word length, word frequency, and stress, as well as random effects such as phoneme identity and word. The models for Dutch and German show many communalities. This is in line with the assumption that similar functional constraints influence reductions in both languages.

Key words: reduction, spontaneous speech, Dutch, German, functional constraints

## Introduction

With the availability of large corpora, it has become clear that spontaneous speech contains many reduced word forms (e.g., yesterday → "yesay"). This gives rise to the questions how words are actually reduced (i.e., why is it "yesay" and not "seday"). Are the reduced form arbitrary "contracts" between speaker and listener, just as the contract that the sound chain "dog" refers to a canine, or are reductions constrained by functional factors?

It has been argued that functional factors influence phonological assimilations (Kohler, 1990; Mitterer, Csépe and Blomert, 2006). Steriade (2001), for instance, argued that directional asymmetries in assimilation are related to the perceptual salience of the assimilated segment. Retroflex stops assimilate mostly in pre-vocalic position, while apical stops assimilate mostly in post-vocalic position. This pattern correlates (negatively) with the salience of place cues for apical and retroflex stops. The formant transitions for retroflex stops are more salient in postvocalic position, while the formant transitions for apical stops are more salient in prevocalic position. Hence, both types of stops assimilate in the position in which their place cues are less salient. This led Steriade to propose the following hypothesis (p. 222): "The likelihood that a lexical representation R will be realized as R' is a function of the perceived similarity between R and R'."

If this constraint also functions for more severe reductions than assimilations, we should find commonalities in reductions over different words and even different languages. To investigate this hypothesis, pronunciations of mid-to-high frequency words with two or three syllables

and six to nine phonemes were investigated in two corpora, a Dutch (CGN, Corpus of Spoken Dutch) and a German (the Kiel Corpus) corpus.

### Method

From the Dutch corpus, 15 words each with 2 or 3 syllables and 6 to 9 phonemes were selected that had a lexical frequency of more than 1 per million and less than 100 per million. For each word, we randomly selected 15 tokens from the corpus which were then phonetically transcribed.

The Kiel Corpus has a more limited vocabulary, so that it was impossible to have a similarly rigorous data gathering approach as for Dutch. Therefore, we selected all words with 2 or 3 syllables and 6 to 9 phonemes from the same frequency range as in the Dutch corpus ( $1/10^6 < \text{frequency} < 100/10^6$ ).

For each phoneme, it was coded whether it was present, weakened, or absent. A linear mixed-effect model with a logistic linking function (Jaeger, in press) was then used to test how well deletion can be predicted by structural (fixed-effect) predictors such as: number of syllables and phonemes, relative position in the word (3d order polynomial), stress, Syllable Part (Onset - Nucleus - Coda), Complexity (is the phoneme part of a consonant cluster) and Word frequency. Additionally phoneme and word were added as random effects. Insignificant factors were deleted by backward elimination.

### Results

Table 1 shows the deletion likelihood in the various major “cells” of the design. Phonemes are more likely to be deleted in words with many phonemes, especially if the word has only two syllables.

Table 1. Deletion likelihood for phonemes in words with different number of phonemes and syllables in English and Dutch.

Language	Number of Syllables	Number of Phonemes			
		6	7	8	9
Dutch	2	0.07	0.14	0.14	0.22
	3	0.09	0.09	0.12	0.14
German	2	0.04	0.07	0.13	--
	3	0.08	0.05	0.08	0.08

The regression analysis (see Table 2) shows that the differences in Table 1 are significant. In both data sets, reduction of a phoneme in a 2-syllable word gets more likely as the word contains more phonemes. There are many other communalities: In both data sets, phonemes are especially likely to be deleted in the middle of words, in the syllable coda, and in unstressed

syllables. The only divergence is that phonemes in consonant clusters (=Predictor Complexity) are only more likely to be reduced in Dutch than phonemes in simple syllable onsets and codas.

Table 2. Partial effects in the regression analysis for both the Dutch and the German samples.

Predictor	Effect in	
	Dutch	German
N(Phonemes) x N(Syllable)	More reduction in long 2-syllable words	More reduction in long 2-syllable words
Position	inverted U-shape	inverted U-shape
Syllable Part	Coda>Nucleus>Onset	Coda>Nucleus>Onset
Stress	Stress inhibits reduction	Stress inhibits reduction
Complexity	More reduction in complex onsets and codas	n.s.
Complexity x Stress	Stronger effect of Complexity with stress	n.s.
Position x Syllable Part	Less reduction for onsets in initial position	Less reduction for onsets in initial position
Lexical Frequency	More reduction in frequent words	More reduction in frequent words
R <sup>2</sup>	0.20	0.27

The analysis of the random effects also shows an interesting pattern. Phonemes that are likely to reduce in one language are also likely to reduce in the other language ( $r = 0.69$ ,  $p < 0.001$ ). In both languages, “s” is very unlikely to be reduced and the alveolar stops /t/ and /d/ as well as the low-amplitude /h/ are likely to be reduced.

## Discussion

The results reveal some expected effects and some unexpected effects. The effect of stress - phonemes in stressed syllables are unlikely to be reduced - is very much expected (Shockey, 2003). However, the fact that phonemes are especially likely to be reduced in word-medial positions is surprising. Traditionally, it is assumed that especially word endings are likely to be reduced. Descriptively, this is true for the current data set as well. Reduction often occur often at the end of words. This effect is however sufficiently explained by the high deletion likelihood for syllable codas and for unstressed syllables. The final segment of a word is often a syllable coda and

(in Dutch and German) often unstressed. This appear to be the causal factors that lead to the deletion of word-final consonants.

Importantly, the strong commonalities in the two data sets are in line with the assumption that reductions are functionally constrained. Although the models only explain a moderate amount of the variance (20%), it is clear that reduced forms do not arise arbitrarily. Some of the effects clearly seem to have a perceptual basis, such as the relatively high likelihood of /h/-deletion. Mielke (2003) already argued that the low amplitude of /h/ - causing it to be perceptually not salient - is the basis for /h/-deletion. Also the (strong) effect of stress can be linked to amplitude: Stressed syllables tend to be longer and louder than unstressed syllables. Reducing segments in stressed syllables would therefore be perceptually salient. Strong conclusions can, however, only be drawn if these results are complemented with independent perceptual evidence, that may or may not converge with the current results.

### **Acknowledgements**

This work was supported by a DFG grant to Holger Mitterer and Mirjam Ernestus in the focus program "Phonological competence" (SPP 1234).

### **References**

- Jaeger, T. F. (in press). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*.
- Kohler, K.J. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. H. A. Marchal (Ed.), *Speech Production and Speech Modelling* (pp. 69-92). Dordrecht: Kluwer.
- Mielke, J. 2003. The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica*, 60, 208-223.
- Mitterer, H., Csépe, V. and Blomert, L. 2006. The role of perceptual integration in the perception of assimilated word forms. *Quarterly Journal of Experimental Psychology*, 59, 1305-1334.
- Shockey, L. 2003. *Sound patterns of spoken English*. Cambridge, MA: Blackwell.
- Steriade, D. 2001. Directional asymmetries in place assimilation: a perceptual account. In E. Hume and K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). New York, NJ: Academic Press.

# Phonological free variation in English: an empirical study

Jose A. Mompean

Department of English Philology, University of Murcia, Spain

## Abstract

This paper presents the results of a corpus-based study of ten words exhibiting phonological free variation in their phonemic or accentual makeup. The study uses data from the News archives of the BBC Learning English website. The rates of use of the variants for each lexical item are given and discussed.

Key words: phonological free variation, free variants, BBC English

## Introduction

Phonological free variation is a well-known phonological phenomenon that occurs when two (or more) phonemes – the free variants – may replace each other in the same position in a word without any change in meaning (e.g. *again* /ə'gen/ or /ə'geɪn/). The phenomenon also applies to words that exhibit different stress patterns (e.g. *controversy* /kən'trɒvəsi/ or /'kɒntrəvɜːsi/) with no change in meaning or grammatical category.

The existence of phonological free variants is caused by different types of factors. These include ongoing sound changes (e.g. /ʃʊə-/ʃɔː/ for *sure* in BrE representing the general replacement of /ʊə/ by /ɔː/ in the system) or phonetic and/or phonological processes such as assimilation, dissimilation, epenthesis or liaison (e.g. /'febjuəri/ for *February* – as well as /'februəri/ – due to dissimilation of the two nearby /r/s). Sociocultural aspects such as speakers' awareness and knowledge or beliefs about the relationship between spelling and pronunciation in the mother tongue or in foreign languages are also a fruitful source of free variation (e.g. /'weɪs'tkəʊt/ for *waistcoat* as well as the former /'wes'tkɪt/-/'wes'tkæt/ in an attempt to follow more closely regular sound/spelling correspondences).

Independently of the causes of phonological free variation, phonological free variants can be related to different variables studied by traditional sociolinguistics. These variables include, for instance, the social/professional group to which the speaker belongs (e.g. /raʊt/ for *route* in BrE army usage vs. /ruːt/, more generally) or the speaker's accent (e.g. AmE /tə'metɪtəʊ/ for *tomato* vs. BrE /tə'mɑːtəʊ/). Another relevant factor is age (e.g. /mɔːl/ for *mall* – shopping centre – preferred by younger speakers to /mæl/, preferred by those born before 1953 as reported by Wells 2008).

Phonological free variation has often been considered as a marginal phenomenon affecting individual (or very small sets of) words. This may be the reason why it has so far received little empirical attention in English. An exception are Gimson's 1969 two corpus-based analyses of spoken speech that aimed at uncovering whether free variation was more common in polysyllables than in monosyllables. More recent studies consist in several surveys of pronunciation preferences carried out by written questionnaire to discover which of the variants for a given set of words the informants preferred. These studies include Wells's three surveys of BrE for the author's *Longman Pronunciation Dictionary* –LPD, for short– (see e.g. Wells 1999, 2008) or Shitara's (1993) survey for AmE.

### **Phonological free variation: an empirical study**

The motivation of surveys of pronunciation preferences such as those conducted for the LPD was to improve on the problem of dictionaries in deciding which variant to prioritise by providing some kind of empirical data regarding the competing pronunciations. However, when the focus of research is on actual usage, analysis of data from spoken language is more appropriate given that the data obtained from the analysis of pronunciation research by written questionnaire and from spoken language may not necessarily coincide.

In the study of phonological free variation, research using real language data was already carried out by Gimson in his 1969 study, although the phonetician did not look at percentages of variation for a given set of lexical items. To our knowledge, this has not been systematically carried out yet for English. As a consequence, in an attempt to look at actual language usage as well as at the productions of individual items exhibiting phonological free variation, a corpus-based study of spoken BrE was planned and conducted. The aim of the study was to uncover rates of usage of different free variants for a given set of lexical items.

### **Method**

#### **Data**

For this study, news archives of the site BBC | Learning English | Words in the News were used (see URL 1). This resource consists of a set of brief newscasts, available both as audio files and as written passages. The archives started in January 1999 and continue up to the present. The newscasts analysed for this study were those from the years 1999 to 2007 (inclusive) read by identified professional newsreaders with an RP – or BBC – accent. The corpus analysed amounts to ca. 180,000 words.

Since the corpus was not produced under the supervision of the author, there is a limitation in the items (as well as their frequency) that can be

studied. As a consequence, words were selected as potential candidates for analysis only if they occurred at least ten times, a criterion that is arbitrary but that somehow guarantees that the word is not too underrepresented in the corpus. With this criterion in mind, ten words were selected for analysis to act as a preliminary report on a deeper study that is currently being carried out with a longer list of words. The items analysed for this paper were six words involving variation in their phonemic structure (*again*, *either*, *neither*, *economic*, *financial*, and *often*) and four words involving variation in stress placement pattern (*controversy*, *kilometre*, *contribute*, and *cigarette*).

### Speakers

The speech of 207 RP newsreaders (145 males, 62 females) was investigated. 41 non-RP newsreaders in the corpus were not considered in this study since we controlled for the variable of the speaker's accent

### Procedure

For the texts produced by the RP newsreaders, the analysis involved identifying the target words in a document containing all the texts pasted from the BBC website, listening to the relevant audio files, and annotating, for each instance of the word studied, the speaker's name and variant used.

### Results and discussion

As the results obtained are shown in Table 1 below. With regards to words with variation in their phonemic makeup, the data show that, in the case of *again*, most instances of the word (85%) contained /e/ in the second syllable as opposed to /eɪ/ (15%). The same rate was obtained for the variant /aɪ-/ word-initially in *either*. The rate is also very high for the variant with /aɪ-/ (75%) in the related word *neither* as opposed to the variant with /i:/ (25%). As far as *economic* is concerned, the rates of usage of the two variants in the first syllable are more similar, with 46% for the option with /i:/ and 54% for the option with /e/. Next, in *financial*, there is a high preference for the option with the diphthong /aɪ/ in the first syllable as opposed to the variants with a monophthong (i.e. /fɪ-/ or /fə-/), analysed jointly (11%). Finally, the variant without a /t/ in *often* is more common (62%) than the option with an epenthetic /t/ (38%), more of a spelling pronunciation.

With regard to words with variation in their stress pattern, *controversy* is pronounced in 75% of the cases on the first syllable (i.e. /'..../) while antepenultimate stress (i.e. /.'.../) is found only in 25% of the cases. In contrast, *kilometre* is pronounced with antepenultimate stress most of the time (73%) as opposed to the variant with stress on the first syllable (27%). However, penultimate stress (i.e. (/.'.../)) is more common in *contribute* (69%) than antepenultimate stress (31%). In the case of *cigarette*,



antepenultimate stress (on the first syllable) obtains a higher percentage (59%) than stress on the last syllable (41%).

Table 1. Words studied, free variants, items found in the corpus & percentage of variant usage.

Word	Variant	Items	%	Word	Variant	Items	%
again	/e/	46	85%	often	/-fn-/	29	62%
	/eɪ/	8	15%		/-ftn-/	18	38%
either	/aɪ-/	23	85%	controversy	/.'.../	9	75%
	/i:-/	4	15%		/.'.../	3	25%
neither	/aɪ-/	9	75%	kilometre	/.'.../	22	73%
	/i:-/	3	25%		/.'.../	8	27%
economic	/ek-/	80	54%	contribute	/.'.../	11	69%
	/i:k-/	69	46%		/.'.../	5	31%
financial	/faɪ-/	88	89%	cigarette	/.'.../	10	59%
	/fɪ,fə/	11	11%		/.'.../	7	41%

## Conclusion

The corpus-based study on phonological free variation in RP described above represents an initial attempt to investigate the rate of occurrence of the different free variants (phonemic and accentual) of specific lexical items in that accent. The data obtained are of interest in themselves but they can also be used as a source of evidence in linguistic theories and/or applications of the latter (e.g. pronunciation teaching). Directions for future research are the eliciting of language data under experimental conditions to obtain evidence on items underrepresented or missing in the corpus, to compare data from different accents or to look into the causes of the variation in specific items.

## References

- Gimson, A. G. 1969. A Note on the Variability of the Phonemic Components of English Words. *Brno Studies in English* 8, 75-79.
- Shitara, Y. 1993. A survey of American pronunciation preferences. *Speech Hearing and Language* 7, 201-232.
- URL 1 [www.bbc.co.uk/worldservice/learningenglish/index.shtml](http://www.bbc.co.uk/worldservice/learningenglish/index.shtml)
- Wells, J.C. 1999. British English pronunciation preferences: a changing scene. *Journal of the International Phonetic Association* 29 (1), 33-50.
- Wells, J.C. 2008. *Longman Pronunciation Dictionary*. Harlow: Pearson Ed.. 3rd ed.

# **Interaction of phonetics, phonology, and sociophonology – illustrated by the vowels of Standard Austrian German.**

Sylvia Moosmüller

Acoustics Research Institute, Austrian Academy of Sciences, Austria

## **Abstract**

Standard Austrian German, which distinguishes eight front vowels, exploits two front constriction locations for distinguishing these vowels, a pre-palatal and a palatal one. However, the pre-palatal location is acoustically unstable, and, consequently, phonologically undesirable. The existence of a sound change which neutralizes the /i/ and /ɪ/ vowels in a first step and shifts the constriction location of the remaining /i/ towards the palatal region in a second step is investigated.

Keywords: phonetics-phonology interface, vowels, sound change

## **Introduction**

All vowels are articulated at a specific constriction location. Stevens (2003) argues that regions of acoustic stability are separated by regions of acoustic instability, and these natural boundaries define the opposition between distinctive features. In Steven's quantal theory (1989), vowel phonemes have an acoustic basis.

Wood (1979, 1982) isolated four contrastive constriction locations for vowels: "along the hard palate, along the soft palate, in the upper pharynx and in the lower pharynx" (1982: 43). Stevens (1989) found three zones – a palatal, a velar, and a pharyngeal zone –, where vowels spectra are relatively insensitive to small displacements in constriction location. For acoustic reasons, the velar and upper pharyngeal constriction are unified by Stevens.

Some languages, however, make use of a further constriction location in the front region. Fant (1970, 2001) describes a pre-palatal constriction location for Swedish and Russian, Wood (1979) for Egypt, Eek and Meister (1994) for Estonian, and Tabain and Perrier (2005) for French. Standard Austrian German (SAG) is a further language which exploits the pre-palatal constriction location (Moosmüller 2007).

Acoustically, the pre-palatal constriction location in /i/ is identified by a substantial rise of F3 due to a switch in the cavity affiliation of F2 and F3 (Stevens 1999). However, this region is acoustically unstable; a widening of the constriction degree or a shortening of the constriction length immediately leads to a lowering of F3.

Phonologically, an acoustically unstable constriction location is undesirable. Therefore, there is a preference to articulate the front vowels at the more stable palatal location. However, in order to shift the four pre-palatal vowels of SAG to the palatal location, it is necessary to reduce the number of front vowels. Consequently, as a pre-requisite to such a shift, the two /i/-vowels and the two /y/-vowels, each pair distinguished by constriction degree, have to be neutralised. This results in a sound change. The following questions arise:

1. Can the pre-palatal constriction location be proved for all speakers?
2. Does a neutralisation of the /i/-vowels and /y/-vowels take place in SAG?
3. Can a shift towards the palatal constriction location be observed?

## Method

Ten speakers (five male, five female), aged from 18 to 84, were asked to read a list of sentences, repeat sentences, and speak spontaneously. The subjects were raised in Vienna with at least one parent raised in Vienna as well<sup>1</sup>. The elder subjects have an academic training, the younger ones are either students or have been to a grammar school.

For the current investigation, all /i/, /ɪ/, and /e/-vowels<sup>2</sup> in stressed position (539 in total) of the spontaneous speech task were segmented manually. F1, F2, and F3 were extracted by means of LPC. A 46 ms long gliding Hanning window was applied with an overlap of 95%. Duration and F0 were measured as well. Depending on the duration of the vowel, the measurement procedure described rendered 20 to 150 measurements per vowel, i.e., the formant frequency contour of the whole vowel was analysed. This method was chosen because vowels, especially when short, often expose no steady state portion.

For statistical analysis, one-tailed t-tests were performed.

## Results

### The pre-palatal constriction location

Looking at the unrounded, constricted vowels, it was observed that the F3/F2 ratio is higher for the pre-palatal vowels than for the palatal vowels. Calculated at the highest point of F3, the F3/F2 ratio of a typical pre-palatal /i/ is  $> 1.4$ . In order to account for a shortening of constriction length in spontaneous speech, the threshold of the ratio is lowered to  $> 1.3$ .

Table 1. Frequency (in %) of F3/F2 ratios of the vowel /i/ in stressed position, spontaneous speech.

Speaker	Year of birth	F3/F2 > 1.3	F3/F2 > 1.4
HK ♂	1923	30.00	3.33
FK ♀	1924	7.69	0.00
BE ♀	1953	57.14	7.14
AS ♂	1956	52.17	21.74
UD ♀	1963	9.09	0.00
FN ♂	1966	69.23	7.69
CL ♀	1976	52.17	13.04
EB ♀	1982	7.69	0.00
FM ♂	1982	5.56	0.00
LH ♂	1989	40.54	8.12

It becomes evident from Table 1 that seven speakers (HK, BE, AS, FN, CL, EB, LX) exploit the pre-palatal constriction location for the vowel /i/. For the remaining speakers (FK, UD, FM), the results remain unclear, since they exhibit no F3/F2 ratios > 1.4. Their results have to be evaluated in conjunction with the results for the vowels /i/ – /e/. Most interestingly, however, no generation-specific dependence of F3/F2 ratios can be found.

It might be argued that the ratio depends on the duration of the vowel, i.e., the longer the vowel, the higher the F3/F2 ratio. However, only one speaker (FM) shows a statistically significant correlation between vowel duration and F3/F2 ratio ( $r = 0.81$ ,  $p < 0.01$ ).

### Neutralisation of the pair /i/ and /ɪ/

Six speakers (HK, UD, FN, CL, EB, FM) neutralise the opposition between /i/ and /ɪ/, whereby /ɪ/ adapts to /i/. FK and BE still distinguish the pair via all three formants, whereas AS distinguishes the pair only via F1, and LH, the youngest speaker, by F2. Despite the fact that these results again do not exactly fit the hypothesis of a generation-specific variation, the assumption of a sound change in progress is nevertheless supported.

### The vowels /i/ and /e/

A shift of /i/ to the palatal constriction location might lead – temporarily – to a neutralisation of /i/ and /e/ with respect to F3, until a new, wider constriction degree is settled for /e/. Only one speaker (FM) shows such a neutralisation, in conjunction with a higher F2 for /i/. The other two speakers who had low F3/F2 ratios (FK, UD) distinguish /i/ and /e/ for F3. Therefore, it can be assumed, that only speaker FM has shifted the vowel /i/ to the palatal location.

## Conclusion

The results presented in this paper strongly point to a current sound change. However, only the first step – the neutralisation of /i/ and /ɪ/ – could be proved. The second step, the shift towards the palatal location, has yet to be accomplished. It may very well be that this step will not take place at all, since phonology is not the only motivation for the neutralisation of /i/ and /ɪ/. The Middle Bavarian dialects, which strongly interact with SAG, distinguish vowel pairs by duration. It is of interest that three out of the six speakers who neutralise /i/ and /ɪ/ for quality, distinguish these vowels by duration. Therefore, the adaptation of /ɪ/ to /i/ might, additionally, be motivated by social factors.

## Notes

- <sup>1</sup> Vienna was chosen for practical reasons, SAG is defined in Moosmüller (1991).
- <sup>2</sup> The /y/, /Y/, and /ø/-vowels have not been analysed, because they make up only 1% of all vowel occurrences (see Moosmüller 2007). The analysis of the vowel /e/ is necessary in order to evaluate the results of /i/.

## References

- Fant, G. 1970. *Acoustic Theory of Speech Production*. The Hague, Mouton.
- Fant, G. 2001. Swedish vowels and a new three-parameter model. *THM-QPSR* 1/2001, 43-49.
- Eek, A. and Meister, E. 1994. Acoustics and perception of Estonian vowel types. *PERILUS XVIII*, 55-90.
- Moosmüller, S. 1991. *Hochsprache und Dialekt in Österreich*. Wien, Böhlau.
- Moosmüller, S. 2007. *Vowels in Standard Austrian German*. Habil.-Schrift, Wien.
- Stevens, K. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.
- Stevens, K. 1999. *Acoustic Phonetics*. Cambridge Mass., MIT Press.
- Stevens, K. 2003. Acoustic and Perceptual Evidence for Universal Phonological Features. In *Proc. of the 15th Intern. Congress on Phonetic Sciences*, vol. 1, 33-38, Barcelona, Spain.
- Tabain, M. and Perrier, P. 2005. Articulation and acoustics of /i/ in preboundary position in French. *Journal of Phonetics* 33, 77-100.
- Wood, S. 1979. A radiographic analysis of constriction location for vowels. *Journal of Phonetics* 7, 25-43.
- Wood, S. 1982. *X-Ray and Model Studies of Vowel Articulation*. Lund Working Papers 23.

# **PENS: a confidence parameter estimating the number of speakers**

Siham Ouamour<sup>1</sup>, Mhania Guerti<sup>2</sup> and Halim Sayoud<sup>1</sup>

<sup>1</sup>USTHB, Institut d'Electronique, BP 32 Bab Ezzouar, Algeria

<sup>2</sup>Ecole Nationale Polytechnique, Algeria

## **Abstract**

Is it possible to know how many speakers are speaking simultaneously in case of speech overlap? If the human brain, creation not yet mastered, manages to do it and even to understand the mixed speech meaning, it is not yet the case for the existing automatic systems. For this task, we propose a new method able to estimate the number of speakers in a mixture of speech signals. The algorithm developed here is based on the computation of the statistical characteristic of the 7th Mel coefficient extracted by spectral analysis from the speech signal. This algorithm using a confidence parameter, which we called PENS, is tested on seven different sets of the ORATOR database, which contain seven multi-speaker files each. Results show that PENS parameter permits us to make a good discrimination, without any ambiguity, between a mono-speaker signal (only one speaker is speaking) and a mixed-speakers signal (several speakers are speaking simultaneously). Moreover, it permits us to estimate, in case of mixed speech signals, the number of speakers with a good precision, especially when the number of speakers is less than four.

## **Introduction**

Often during discussions, debates and confrontations, when several speakers share a discussion, we are in presence of simultaneous speech mixture of several speakers, due to the intervention of these speakers in the same time, during the discussion: Takayuki Arai (Arai 2003). Thus, the speech signal will contain some zones of speech overlap: F. Asano (Asano 2007).

Such cases often arise with female speakers: women have a multi-task behavior (Changingminds.org) which permits them to speak and understand in such conditions, although that case may also arise with male speakers, often during hot debates between adversary presenting opposite ideas, such as political debates for example. Moreover, those speech overlaps may characterize specifically one language more than another: for instance, in certain regions of Italy (Changingminds.org) people are known by the fact to begin to speak even before the other interlocutor has finished his sentence.

However, in audio document indexing by speakers, those overlap zones remain difficult to index, since we cannot attribute them to a single speaker alone. So it is interesting to know these zones locations even before applying the indexing system. For that reason, we have developed a new algorithm able to discriminate between a mono speaker speech signal and multi-

speaker speech signal containing several speakers speaking in the same time. This algorithm has many applications: it can be applied, for instance, to an audio document, just before the indexing phase in order to avoid and eliminate the segments presenting such ambiguities.

In the rest of the paper, we will present our new algorithm and show the experimental results. We will conclude at the end of the paper by giving some discussions on the results.

### Approach description

This research work deals with the estimation of the number of speakers in a speech mixture: Takayuki Arai (Arai 2003). Our approach is based on the statistical characteristics of the 7<sup>th</sup> Mel filter ( $mel_7$ ): H. S. Lee and A.C. TSOI (Lee 1995), as described in table 1.

Table 1: Spectral characteristics of the 7<sup>th</sup> Mel filter

	Cut-off frequency at 0%	Cut-off frequency at 50%
Fmin	1.3750 kHz	1.6125 kHz
Fmax	2.3750 kHz	2.0813 kHz
Median Freq.	1.8375 kHz	1.8375 kHz

In reality, the discovery of a confidence parameter, estimating the number of voices in a speech signal, was found after several experimental trials, but none of the other tested parameters was interesting: only one was pertinent. We called this pertinent parameter: ‘Parameter Estimating the Number of Speakers’ (PENS). This parameter is given by:

$$PENS = \overline{mel_7} - \sqrt{\text{var}(mel_7)} \quad (1)$$

where  $\overline{mel_7}$  represents the mean of the 7<sup>th</sup> Mel filter ( $mel_7$ ).

As explained previously, many experimental attempts were made, but we kept only the parameter having a strong impact on the speakers number.

### Results and interpretation

We recall that the principal objective, expected by this research work, is the estimation of the number of speakers speaking simultaneously during an interview or multi-conference.

For that reason, we have tested the PENS parameter on seven different subsets of the ORATOR database (Ouast 2002), containing seven speech files of 8 seconds each. The experiments are divided into 2 series:

### First evaluation: estimation of the number of speakers

This evaluation experiment consists in trying to estimate the number of speakers speaking simultaneously in a sequence of a speech file. Results of estimation are presented in figure 1.

We notice on figure 1 that we can easily know if the speech file contains the speech of only one speaker or a speech mixture of two speakers or more.

So, we can easily deduce if only one speaker is speaking or if it is a speech overlap of different speakers; this estimation is then accurate with a precision of 100% since the separation range between the files of a single speaker and the other cases is considerable.

For the files containing three mixed speakers, the estimation error is 14.3%. This error increases if the number of speakers exceeds 4 speakers (figure 1).

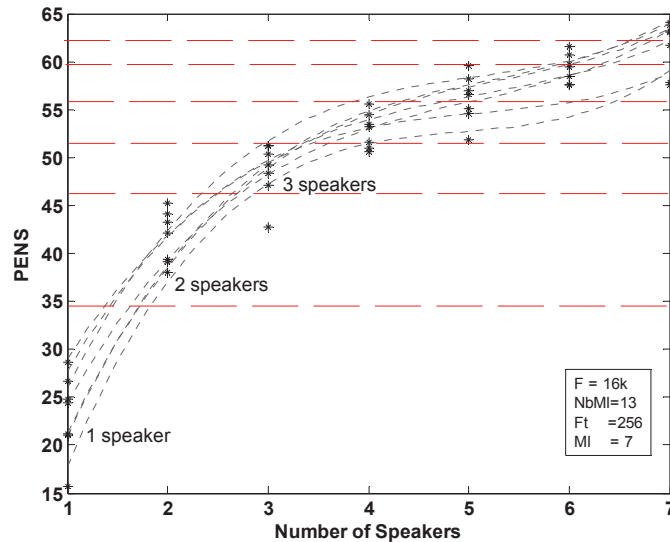


Figure 1: PENS versus the number of speakers.

### Second evaluation: test of discrimination

This test consists in making a discrimination between two speech signals according to the speakers number. Results of discrimination, presented in table 2, show that the discrimination between an audio document containing the speech of a unique speaker and another one containing the speech of two speakers can be made without any ambiguity and without any error. We get the same result if the difference between the speakers number is more than one (e.g. between 2 and 4 speakers or between 3 and 5 speakers, etc...).



Table 2: Discrimination according to the number of speakers in %.

Discrimination	Good discrimination	Error of discrimination
1 and 2 speakers	100	0
1 and several speakers	100	0
2 and 3 speakers	92.9	7.1
2 and 4 speakers	100	0
3 and 4 speakers	85.7	14.3
3 and 5 speakers	100	0
4 and 5 speakers	78.6	21.4
4 and 6 speakers	100	0

### Conclusion and interpretation

The objective of this work is to find a confidence parameter allowing us, in one hand to distinguish a mono-speaker speech segment from a multi-speaker speech segment, and in the other hand, to estimate the number of speakers sharing a discussion simultaneously with the lowest error possible. Experimental results show that the use of the new confidence parameter PENS gets an interesting precision. The discrimination between single and multi-speaker segments has an error of 0%, and the estimation of the number of speakers, talking in same time, has an error of 0% for a unique speaker, an error of 0% for two speakers and an error of 14,3% for three speakers. Over four speakers, the estimation becomes less accurate.

Finally, we consider that this research domain remains little explored even though the problems of speech overlap, encountered in practice, are very restrictive in speech recognition or audio indexing.

### References

- Arai, T. 2003. Estimating Number of Speakers by the Modulation Characteristics of Speech. ICASSP, 197-200.
- Asano, F., Yamamoto, K., Ogata, J., Yamada, M. and Nakamura, M. 2007. Detection and separation of speech events in meeting recordings using a microphone array. EURASIP Journal on Audio, Speech, and Music. Volume 2007, ID 27616.
- Overlapping speech. [http://changingminds.org/techniques/conversation/interrupting/overlap\\_speech.htm](http://changingminds.org/techniques/conversation/interrupting/overlap_speech.htm)
- Lee, H.S. and A.C. TSOI. 1995. Application of multi-layer perceptron in estimating speech / noise characteristics for speech recognition in noisy environment. Speech Com. 17, 59-76.
- Quast H. 2002. Automatic recognition of nonverbal speech. An Approach to model the perception of para- and extraling. Vocal Commun. with Neural Net. Mach. Per. Inst. for Neural Comput. UC San Diego, June 28 2002.

# Phoneme classification using the Hartley Phase Spectrum

Ioannis Paraskevas and Maria Rangoussi

Department of Electronics, Technological Educational Institute (T.E.I.) of Piraeus,  
Greece

## Abstract

The phase of a signal conveys critical information for signal interpretation. However, signal phase extraction is not a straightforward process, mainly due to the discontinuities appearing in the phase spectrum. Previously, it was shown that for the application of audio classification, in case features of the processed phase spectrum are combined with magnitude related features they provide higher classification score compared to the case where features are extracted only from the signals' magnitude content. Moreover, it was shown that the Hartley Phase Spectrum encapsulates the phase content of the signals more efficiently compared to its Fourier counterpart. In this work, the importance of phase and the properties of the Hartley Phase Spectrum are demonstrated for the application of phoneme classification.

Key words: phoneme classification, Fourier Phase Spectrum, Hartley Phase Spectrum

## Introduction

In this work we introduce the use of the Hartley Phase Spectrum (HPS) as an alternative to the standard Fourier Phase Spectrum (FPS) for the classification of speech signals. Accurate phase information extraction is critical for the success of the subsequent speech processing steps; yet difficulties in processing the phase spectrum of speech signals (Alsteris 2006) have led researchers to focus their investigation to magnitude based features only.

In speech signals (and other audio signals), the shape of the FPS is characterized by rapid changes, appearing as discontinuities, which are caused i) due to the use of the inverse tangent function, which is a highly discontinuous function ('extrinsic' discontinuities) (Tribolet 1977) and ii) due to properties of the signal itself ('intrinsic' discontinuities) (Al-Nashi 1989). The difficulty in processing the phase spectrum of signals, due to the aforementioned discontinuities, led the researchers to extract spectral features mainly from the magnitude spectrum and hence, ignore the phase related content of the signals (Paraskevas 2004).

In the FPS, the heuristic approach for the compensation of the 'extrinsic' discontinuities using the 'unwrapping' algorithm introduces ambiguities

which are more severe in the presence of even the lower possible noise level. The advantage of the HPS, compared to its Fourier counterpart, is that it does not convey ‘extrinsic’ discontinuities whereas its ‘intrinsic’ discontinuities can be easily compensated or removed when needed. The HPS, due to its structure, is less affected by the presence of noise; the noise immunity of the HPS, compared to its Fourier counterpart, is justified via the shapes of the respective probability density functions (Paraskevas 2008). The superiority of the HPS and its corresponding cepstrum, compared to its Fourier counterparts, has already been shown in signal analysis and pattern recognition (Paraskevas 2006 and 2007).

In this work, features are extracted from the processed phase spectrogram and are combined with features extracted from the magnitude spectrogram of the signals, for the application of phoneme classification. Specifically, we aim to classify four consonants namely: /h/ e.g. in “hill”, /s/ e.g. in “sister”, /f/ e.g. in “fat” and /th/ e.g. in “thin” (TIMIT® Acoustic-Phonetic Continuous Speech Corpus), based on their spectral characteristics. Similarly to our previous work - which involved audio signals (Paraskevas 2006) - the experimental results show that the classification score obtained in the case where the features extracted from the magnitude spectrograms are combined with the features extracted from the processed phase spectrograms of the phonemes is higher, compared to the case where the features employed are extracted only from their magnitude spectrograms.

### **Experimental set-up**

The feature extraction step of the pattern recognition process includes calculation of statistical features from each spectrogram. Thus, each recording is represented by an  $[8 \times 1]$  - sized feature vector. The statistical features employed here are chosen empirically as the most representative descriptive statistics: the variance, the skewness, the kurtosis, the entropy, the inter-quartile range, the range, the median and the mean absolute deviation.

In steps, each recording is divided into equal-length frames (256 samples) with zero-padding of the last frame if necessary, windowed with a Hanning window and transformed to the frequency domain. The transformed frames are placed row-wise in a matrix and hence the spectrograms are formed (Fourier and Hartley Phase Spectrograms, Fourier and Hartley Magnitude Spectrograms and Hartley Transform (Bracewell 1986) Spectrogram). Each of the spectrograms has to be presented to the classifier with its dimensionality reduced. Hence, the statistical features are calculated from each spectrogram for each recording, in order to compress the information into a compact feature vector. Then, from the aforementioned spectrograms, three were selected based on their recognition performance, namely: i) the

Fourier Magnitude (FM) Spectrogram, ii) the Hartley Phase Spectrogram (HPS) via the Discrete-Time Hartley Transform (DTHT) and iii) the Hartley Magnitude (HM) Spectrogram - all three spectrograms convey distinct information.

The classification decision for each recording is based on the ‘majority vote’ classification concept. For each recording, the statistical features are calculated for each of the three aforementioned spectrograms. Consequently, three feature vectors are formed for each recording. A phoneme is classified to a certain class based on the minimum distance between the test and the target vectors where the classifier used for each of the three independent ‘experts’ is the Mahalanobis (Mahalanobis 1936) and (Webb 2002). In case two or three out of the three independent ‘experts’ agree, then a recording is classified to this class. In case of tie, the decision is taken based on the Fourier Magnitude independent ‘expert’. Each class of phonemes - /h/, /s/, /f/ and /th/ - consists of ten recordings where seven of them are used as training data and the rest as test data.

### Experimental results

A series of two-class experiments between the aforementioned speech signals are carried out - Table 1 presents the correct classification scores, averaged across all six two-class experiments.

Table 1. Correct classification scores (%).

Feature	FM	FM, HPS via the DTHT & HM
Average scores (4 classes)	52.8	75.0

The second column of the table presents the correct classification score in the case where the features are extracted only from the Fourier Magnitude spectrogram and the third column presents the correct classification score of the combined scheme (Fourier Magnitude, HPS via the DTHT and Hartley Magnitude spectrograms). Similarly to the experiments involved the audio signals (Paraskevas 2006), the classification score is increased (22.2%) in the case where the phase-related features are also employed, compared to the case where features are extracted only from the Fourier Magnitude content of the speech signals. Moreover, as in (Paraskevas 2006), the experimental results show that the classification rate is higher in the case where i) the statistical features are extracted from the Hartley rather than the Fourier Phase spectrogram, ii) the discontinuities are compensated, iii) the ‘difference’ is used when estimating the phase spectrum.

## Conclusions

This work - focused on the feature extraction stage of the pattern recognition process - shows the importance of phase as a feature extraction tool for phoneme classification. The experiments involving speech signals show that the proposed phase-related features, when combined with conventional features based on the magnitude spectrum, form a significantly more discriminative feature set. Moreover, the experimental results indicate that the features extracted from the Hartley Phase Spectrum, present the information content of the signals to the classifier in an improved manner, as compared to the Fourier Phase case. As a conclusion, features extracted from the Hartley Phase Spectrum, if employed along with conventional magnitude related features, provide a significantly improved feature set for frequency-domain statistical classification.

## References

- Al-Nashi, H. 1989. Phase Unwrapping of Digital Signals. *IEEE Transactions on Acoustics, Speech and Audio Processing* 37 (11), 1693–1702.
- Alsteris, L.D., and Paliwal, K.K. 2006. Further intelligibility results from human listening tests using the short-time phase spectrum. *Speech Communication*, 48, 727-736.
- Bracewell, R. N. 1986. The Fourier Transform and Its Applications. In 2nd ed. chap. 19, McGraw-Hill, New York.
- Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India*, vol. 12, pp. 49-55.
- Paraskevas, I., and Chilton, E. 2004. Combination of Magnitude and Phase Statistical Features for Audio Classification. *Acoustics Research Letters Online*, vol. 5, Issue 3.
- Paraskevas, I., Chilton, E., and Rangoussi, M. 2006. Audio Classification Using Features Derived From The Hartley Transform. *Proc. of the 13th Int. Conference on Systems, Signals and Image Processing*, Budapest, Hungary, pp. 309-312.
- Paraskevas, I., and Rangoussi, M. 2007. The Hartley Phase Cepstrum as a Tool for Signal Analysis. *Advances in Nonlinear Speech Processing*, Springer Lecture Notes in Computer Science, vol. 4885, pp. 204 - 212.
- Paraskevas, I., and Rangoussi, M. 2008. The Hartley Phase Spectrum as a noise-robust feature in speech analysis. To appear in *Proc. of the ISCA Tutorial and Research Workshop (ITRW) on Speech Analysis and Processing for Knowledge Discovery*. Aalborg, Denmark.
- Tribolet, J. 1977. A new phase unwrapping algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 2, pp. 170 – 177.
- Webb, A. 2002. *Statistical Pattern Recognition*. In 2nd ed., John Wiley and Sons, England.

# **The influence of music education and training on SLA**

Barbara Pastuszek-Lipińska

School of English, Adam Mickiewicz University, Poland

## **Abstract**

Although music education is one of the human activities that requires the integration of all human senses and the involvement of all cognitive processes: sensory, perceptual, and cognitive learning, memory, emotion, and auditory and motor processes, music has tended to figure only marginally in an approach to second-language acquisition. To explore the extent to which music education influences second-language acquisition musicians and non-musicians were asked to reproduce thrice-repeated sentences in six languages. Musicians outperformed non-musicians in the study. From the results, it appears that the influence of musical expertise extends beyond music processing to speech processing.

Key words: music education, second language acquisition

## **Introduction**

Language and music are universal among human cultures. Both are conveyed by sequences of sounds organized in time, and the temporal or rhythmic aspects are highly important features of both domains. Both domains also involve organized acoustics signals that are used in interpersonal communication, and both involve complex cognitive and motor processes. Thus, it is not surprising that the two domains and the relationship between them have attracted interest of researchers from a variety of disciplines (Magne et al. 2003, Thompson et al. 2004, Jentschke et al. 2005, Lahav et al. 2005, Schellenberg 2005, Magne et al. 2006).

Although the similarities and differences between the two systems have attracted the attention of scientists and researchers from different disciplines for centuries, most of the analyses were conducted during the last three decades, when the issues emerged with a new impetus due to new research methods and technical possibilities. Since then, a fair number of researchers from different domains have examined the issues from different points of view.

## **Research design**

### **Research corpus**

82 word sequences in 6 languages (English: American (15), British English (14), Belgian Dutch (10), French (10), Italian (10), Spanish: European (6) and South American (4) and Japanese (10)) were synthesized for the corpus. The ScanSoft® RealSpeak™ application was used for this purpose.

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.

Languages were chosen according to their typological classification; among them there were stimuli that included stress-timed, syllable-timed, and morae-timed languages. Among the sequences were questions, statements, and orders.

### **Participants**

In the reported study participated 106 (53 musicians and 53 non-musicians) non-paid volunteers, Polish native speakers. All subjects were aged from 15 to 69 years, with a mean age of 32 (median 28).

### **Research procedure**

Participants were asked to reproduce as accurately as they could foreign language sentences - synthetic stimuli after three repetitions with taking into account both segmental (vowels and consonants) and suprasegmental (intonation, rhythm, stress, and rate) features. All subjects' productions were recorded, and, with data gathered from the speakers through special questionnaires, were examined with a battery of tests and analyses.

Additionally, the musical skills of the participating non-musicians were tested (Pastuszek-Lipińska 2003). Data were analysed through several different experiments two of them are shortly presented below.

#### **Experiment 1 – General auditory analysis**

All recordings were rated by the author by an impressionistic auditory analysis. In the first round of data analysis, the scoring procedure was based on a general review and observation of whether all speakers responded to the stimuli and were able to repeat the speech material in the given time and with appropriate accuracy.

#### **Experiment 2 – Listening test 2 – web-based cross-linguistic listening test**

To observe how participants were perceived by native speakers of all languages involved in the study, a web-based experiment with a panel of native speakers of involved languages was designed and conducted.

The group of raters consisted of twenty four native speakers of American English, two native speakers of Belgian Dutch, four native speakers of British English, fifteen native speakers of French, eight native speakers of Italian, three native speakers of Japanese, and nine native speakers of Spanish.

Judges rated subjects' productions in only their mother tongue using the Visual Analogue Scale.

## Research results

### Experiment 1 – results

Each participant encountered some problems with at least one sentence. However, 65.53% of musicians' and 46.55% of non-musicians' productions were rated as correct. So the general analysis revealed better performance of musicians comparing with non-musicians.

### Experiment 2 – results

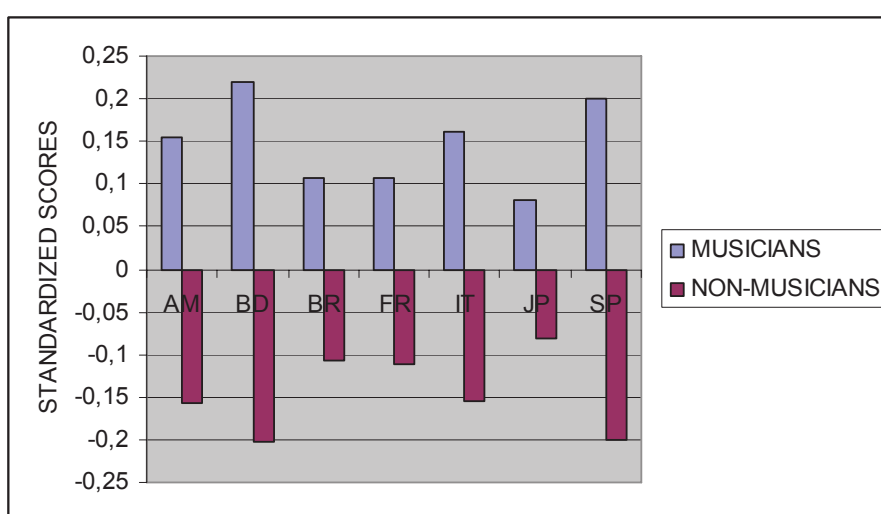


Figure 1. Mean standardized scores obtained by two groups of examinees.

The graphs above present the standardized scores given by the native speakers; it can be seen that in all examined sentences, musicians obtained higher scores (above the average at 0 level), while non-musicians obtained lower scores (below the average).

## Conclusions

Results confirm the hypothesis that musicians would perceive and produce speech sequences better than non-musicians. Moreover, this trend was observed in syllable, stress, and morae-timed languages.

It seems to be clear that even without obvious correlations between variables, the results provide evidence that music education has a significant influence on the acquisition of a second language and more specifically pronunciation.

The results revealed that music education exerted a measurable impact on speech perception and production. Musicians outperformed non-musicians in



the study. Therefore, the superior performance of the musicians in the task may be interpreted as evidence that music education is an enabling factor in the successful acquisition of a second language.

## References

- Jentschke, S., Koelsch, S. and Friederici, A.D. 2005. Investigating the relationship of music and language in children. Influences of musical training and language impairment. In Avanzini, G., Lopez, L., Koelsch, S. and Majno, M. (eds.), *Annals of the New York Academy of Sciences*, vol. 1060, 231–242, New York, USA.
- Lahav, A., Boulanger, A., Schlaug, G. and Saltzman, E. 2005. The power of listening. Auditory-motor interactions in musical training. In Avanzini, G., Lopez, L., Koelsch, S. and Majno, M. (eds.), *Annals of the New York Academy of Sciences*, vol. 1060, 189–194, New York, USA.
- Magne, C., Schön, D. and Besson, M. 2003. Prosodic and melodic processing in adults and children: behavioral and electrophysiologic approaches. In Avanzini, G., Faienza, C., Minciacchi, D., Lopez, L., and Majno, M. (eds.), *Annals of the New York Academy of Sciences*, vol. 999, 461–476, New York, USA.
- Magne, C., Schön, D. and Besson, M. 2006. Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience* 18(2), 199–211.
- Pastuszek-Lipińska, B. 2003. Unpublished test of musical abilities.
- Schellenberg, E.G. 2005. Music and cognitive abilities. *Current Directions in Psychological Science* 14, 322–325.
- Thompson, W.F., Schellenberg, E.G. and Husain, G. 2003. Perceiving prosody in speech effects of music lessons. In Avanzini, G., Faienza, C., Minciacchi, D., Lopez, L., and Majno, M. (eds.), *Annals of the New York Academy of Sciences*, vol. 999, 530–532, New York, USA.

# **Rhythmic analysis and quantitative measures: the essence of rhythm as temporal patterning**

Michela Russo

University of Paris 8/UMR 7023-C.N.R.S., France

## **Abstract**

A comparison of Standard Italian and Southern Italian dialects is carried out using acoustic measures related to structural properties: Vowel-interval and consonantal inter-vowel-interval durations are used to obtain rhythmic measures based on a number of approach. Italian is usually considered to be ‘syllable-timed’. The structural features found in the dialects offer support for a divergence from the traditional assumption of syllable-timing. Rhythm measures are calculated according to Ramus et al. 1999, Grabe-Low 2002, Barry et al. 2003, Russo-Barry to appear. In agreement with predictions derived from phonological observation, the results of the measures show a ‘rhythm plot’ in which the ‘Pairwise Variability Indices (PVIs)’ place the Italian dialect speakers nearer to the stress-timed’ languages than traditional typology statements would lead one to expect .

Key words: language typology, Italian dialects, rhythmic values

## **What differences are there between the languages?**

In traditional views of rhythmic typology the rhythmic classification of a language was considered a given ‘primitive’, an inherent property of the language. In contrast, more recent studies see it as an emergent property, a product of both the phonotactics of the language and phonetic processes in production (Dauer 1987). Syllable complexity, possibly vocalic and consonantal length distinctions, stress-dependent vowel reduction and the propensity for phonetic vocalic and consonantal reduction processes (‘schwa-isation’, weakening and elision, etc.) during speech are considered contributory factors to the rhythm of an utterance, and via this to the general rhythmic impression of a language. Critically, rhythm not only becomes measurable, as a speech phenomenon rather than an inherent language property, but it also necessarily becomes a continuous rather than a categorical property.

Italian is usually considered a ‘syllable-timed’ language: it has a relatively simple (CV-dominated) basic syllable structure, no phonological vowel length opposition and no phonological vowel reduction. On the other hand, it has a consonantal length distinction, and pronounced allophonic tonic-vowel lengthening. Phonetic and phonological evidence supports the interpretation of Southern Italian dialects (Ischia, Capri, and the dialect of Pozzuoli, near Naples) as a stress-timed language. Distributional

observations and durational measurements of tauto-and heterosyllabic VC sequences show that make a strictly syllable-timed rhythmic structure untenable. The structural features found in the dialects that offer support for a divergence from this traditional assumption are: a) Long vowels or long diphthongs in closed syllables; b) Neutralisation of vowel timbre in unstressed syllable; c) Loss of unstressed vowels. Although much of the dialect observations place the stress-timing evidence at the systemic rather than the realisational level, the non-systemic, ‘performance’ evidence points in the same direction.

### Rhythm measures

Rhythm measures are calculated according to Ramus et al. 1999, Grabe-Low 2002, Barry et al. 2003, Russo-Barry to appear. The Ramus measures are (i) the proportion of vowels in the interpause stretches ips (%V), (ii) the standard deviation of the Vowel duration in the ips ( $\Delta V$ ) and (iii) the standard deviation of the intervocalic consonantal interval ( $\Delta C$ ). The Grabe and Low measures correspond in essence to the Ramus variability measures, but are calculated in pairwise steps through the ips rather than globally across the ips. They are therefore called ‘Pairwise Variability Indices’ (PVIs).

(i) Non-normalized consonantal PVI:

$$^r PVI = \left[ \sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right]$$

(ii) Normalized vowel PVI (for vowels to correct for tempo fluctuations):

$$^n PVI = 100 \times \left[ \sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{(d_k + d_{k+1})/2} / (m-1) \right]$$

The difference (i) between consecutive vowels and (ii) between consecutive intervocalic intervals are averaged over the ips, giving a vocalic and consonantal variability measure. In the case of the vowel intervals, the difference is related to the sum of the two vowels. This ‘normalisation’ is claimed to be necessary (and possible) for the vowel intervals in order to counteract shifts in tempo because vowels vary more than consonants with tempo, and there is never more than one vowel in a vowel interval.

Corpora had been segmented and labelled, providing the segmental identities and durations which form the basis of the rhythm measures. Pauses, hesitations and other interruptions had also been annotated, so it was possible to identify prosodically uninterrupted ‘inter-pause stretches’ (ips):

Table 1. Ramus (%V,  $\Delta V$ ,  $\Delta C$ ) and Grabe-Low (PVI-V, PVI-C) rhythm measures for the Naples, Pisa and German speaker groups.

Lang.group	%V	$\Delta V$	$\Delta C$	PVI-V	PVI-C
Naples	54.2	59.8	51.5	39.2	56.4
Pisa	55.1	65.4	51.0	43.0	58.9
German	42.0	42.2	64.5	55.0	65.1

The Italian here consists of semi-spontaneous speech recordings from the AVIP/API regional database (*Archivio Varietà Italiano Parlato*, <ftp://ftp.cirass.unina.it> – Map-Task dialogues) and the German is the Kiel Corpora of read and spontaneous speech (IPDS 1994, 1995).

As showed from the average group values given in Table 1, what is most striking about the measures is the much higher vowel variability of the Italian speakers. They are further from the Spanish values found in previous studies even than English and German. These may lie in the amount of speech material, in the type of speech material, or in the selection of speakers. Furthermore, we give here measures for 10 fluent sections of spontaneous speech from the Neapolitan dialect of Ischia (Forio, a total of 33.62 sec). The average PVI scores for one speaker is: Raw PVI (Consonant interval) 52.52, Normalized PVI (Vowel interval) 55.98. The average percentage vocalic interval in the utterances is 54.9%. Compared to our earlier data and to data in the literature these measures are different to some extent, but only in the consonantal measure: the %V value of 54% is clearly equivalent to the Italian values we got for Bari, Pisa and Napoli and much higher than any of the German or Bulgarian values (they never reached 50%, even at the fastest tempo, Barry et al. 2003). Our values for Bari, Pisa and Napoli are (from the AVIP/API corpus): Raw PVI-C Bari 61.6, Pisa 58.9, Napoli 56.4; Norm V-PVI Bari 41.6, Pisa 43.0, Napoli 39.2. So the speaker consonant variability is lower (there is no support for ‘stress timing’), but the vowel variability is considerably higher (pushing the vowels away from syllable timing). For comparison, our German values were: Raw PVI-C - German 68; Norm PVI-V - German 55. The values from Grabe-Low 2002 are: German 59.7/55.3, English 64.1/57.2, French 50.4/43.5, Spanish 57.7/29.7. In terms of text-dependent variation in the values, Grabe-Low 2002 publish the following values for 3 different parts of their material: PVI-C German 52.1/57.0/55.9; English 65.6/65.0/54.4; French 49.3/49.7/44.3; Spanish 60.3/56.9/54.7; PVI-V German 57.6/65.3/58.7; English 55.2/53.6/56.1; French 39.4/38.7/42.0; Spanish 26.4/27.7/26.0. We compare these values with the variation over 10 utterances of the dialectal speaker’s. Raw PVI-C: 56.4, 37.1, 44.6, 36.7, 54.6, 71.7, 55.3, 44.8, 49.7, 74.3; Norm PVI-V: 45.2, 52.7, 63.7, 64.5, 51.3, 55.6, 60.2, 65.7, 51.2, 49.7. The values

above show a ‘rhythm plot’ in which the PVI-V groups the dialectal speaker with ‘stress-timed’ languages against the traditional typology expectation.

### **Acknowledgements**

Grateful thanks to William John Barry, Caren Brinckmann, Bistra Andreeva and Anja Moos for processing the labelled data used for the analysis.

### **References**

- Barry, W.J. and Russo, M. 2003. Measuring rhythm. Is it separable from speech rate?. In Mettouchi, A. and Ferré, G. (eds.), *AAI Workshop, Prosodic Interfaces*, 15-20. Nantes, Université de Nantes, UFR Lettres et Langage, AAI.
- Barry, W.J. and Russo, M. 2004. Isocronia oggettiva o soggettiva? Relazioni tra tempo articolatorio e quantificazione ritmica. In Albano Leoni, F., Cutugno, F., Pettorino, M. and Savy, R. (eds.), *Il parlato Italiano*, cdrom A02. Napoli, D’Auria.
- Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S. and Kostadinova, T. 2003. Do Rhythm Measures Tell us Anything about Language Type?. In Solé, M., Recasens, D. and Romero, J. (eds.), *Proc. of the 15th Intern. Congress of Phonetic Sciences*, 2693-2696. Barcelona: Causal Productions Pty Ltd.
- Dauer, M.R. 1987. Phonetic and phonological components of language rhythm. In *Proc. Of 11th Intern. Congress of Phonetic Sciences*. Tallinn, 447-450. Estonia: U.S.S.R. / Academy of Science of the Estonian S.S.R., vol. 5.
- Grabe, E. and Low, E.L. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. In Gussenhoven, C. and Warner, N. (eds.), *Papers in Laboratory Phonology VII*, 515-546. The Hague: Mouton de Gruyter.
- IPDS 1994. The Kiel Corpus of Read Speech, vol. 1, CD-ROM 1. IPDS 1994-1997. The Kiel Corpus of Spontaneous Speech, vol. 1-3, CD-ROM 2-4. Kiel: Institut für Phonetik und digitale Sprachverarbeitung.
- Ramus, F., Nespor, M. and Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
- Russo, M. and Barry, W.J. 2004. In che misura l’italiano è ‘iso-sillabico’? Una comparazione quantitativa tra l’italiano e il tedesco. In D’Achille, P. (ed.), *Generi, Architetture e forme testuali*, 387-401. Firenze, Cesati.
- Russo, M. and Barry, W.J. 2008. Isochrony reconsidered. Objectifying relations between Rhythm Measures and Speech Tempo. In Barbosa P., Madureira, S. and Reis C., *Proc. of 4th Conference on Speech Prosody 2008*, Campinas, May 6-9, 2008, Brazil, cd rom.

# Compensatory lengthening in Persian: the timing of non-modal phonation

Vahid Sadeghi

Department of English Language, Imam Khomeini International University, Iran

## Abstract

Persian has CVGC (or CVCG) sequences (G, a glottal consonant; /h/ or /ʔ/) which become reduced in certain occurrences, with the perceptual effect of the loss of the glottal consonant. The purpose of this study is to provide an acoustic description of the sequences in reduced forms. A production study examined three acoustic measurements of phonation types: H1-H2, H1-F1, and f0. The measurements were made at 15 ms time intervals throughout the vowel to determine the time course of phonation effect. Results indicate that G is realized as breathiness in CVhC (or CVCh) and laryngealization in CVʔC (or CVCʔ), and that such non-modal voice qualities spread onto half of the portion of the preceding vowels.

Key words: Spectral tilt, compensatory lengthening, voice quality variation

## Introduction

It has long been suggested that Persian exhibits compensatory lengthening (CL), which is triggered by the deletion of glottal consonants in coda position in informal speech (Shademan 2003). This implies that Persian could possibly contain surface minimal pairs, such as /bæʔd/ → [bæ:d] “later” and /bæd/ → [bad] “bad” that would contrast only with respect to vowel length. Results of recent studies, however, indicate that vowel offset f0 and spectral tilt measurements for CVGC tokens changed significantly relative to CVC, suggesting CL might better be interpreted as a quantitative variation of glottal gesture (Bijankhan 2000).

The present paper attempts to explore the time course of voice quality variation in CL data. Measurements of the time course of variation in phonation in the data indicate that non-modal phonation in CVGC data is not localized to vowel offset, as was already assumed; rather, glottal consonants extend their non-modal quality over half of the portion of the preceding vowel. Thus, it is suggested that CL involves a sequencing of modal and non-modal phonation. The portion characterized by modal quality corresponds to the entire length of the vowel preceding glottal consonants, and the second portion which is associated with non-modal voice quality and has a shorter duration relative to the modal part corresponds to the entire length of the coda glottals, realized with magnitude reduction of glottal gesture. Thus, the additional length associated with CL data could result

from the laryngeal variation of coda glottals that adds an interval of non-modal vocalic gesture to the preceding vowel.

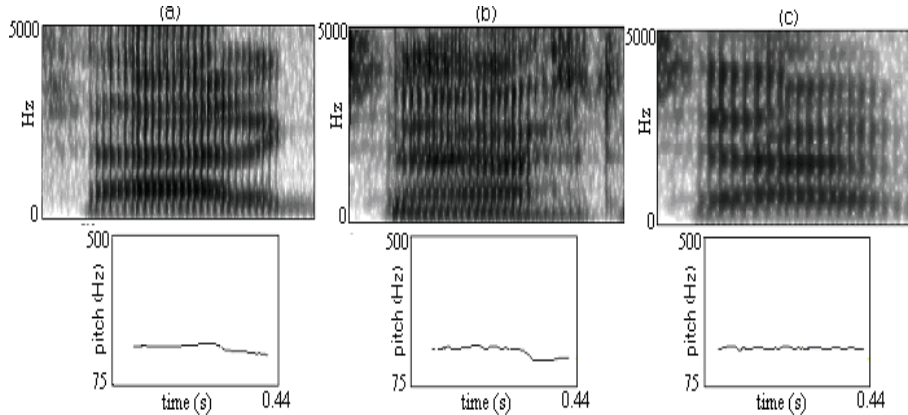


Figure 1. The spectrograms and  $f_0$  contours for the words [shar?] (a), [sharh] (b) and [shar] (c), by the same speaker.

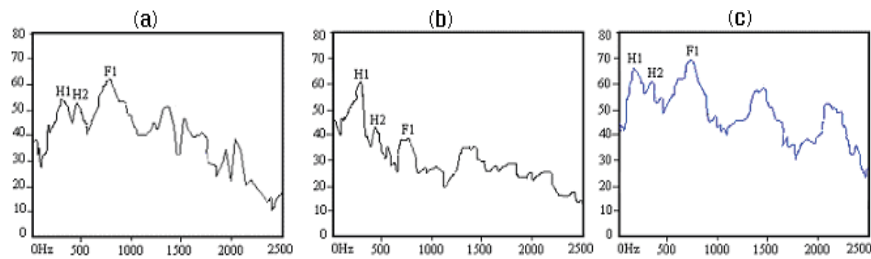


Figure 2. FFT spectra for the vowel [a] in [shar?] (a), [sharh] (b) and [shar] (c) sampled at 45 ms to the vowel offset.

### Phonetics of phonation types

Phonation types can be quantified through a number of phonetic parameters. These include fundamental frequency and spectral tilt, among others. Non-modal-phonation types are commonly associated with lowering of fundamental frequency (Gerfen and Baker 2005), (Gordon and Ladefoged 2001). Spectral tilt is also known to differentiate phonation types in a number of languages including Jalapa Mazatec (Blankenship 2002), Tagalog (Blankenship 2002), and !Xoo (Gordon and Ladefoged 2001). It can be quantified by comparing the amplitude of the first harmonic to that of the second and higher frequency harmonics. It is assumed that spectral differences yield positive values for breathy vowels which have the greatest drop off in energy as frequency increases, and negative values for creaky vowels which display the largest boost in energy at higher frequencies.



Displayed in figure 1 are the spectrograms, f0 contours for the words [ʃarʔ] ‘religion’ (a), [ʃarh] ‘explanation’ (b) and [ʃar] ‘evil’ (c), produced by the same speaker. In [ʃarʔ], though there is little obvious evidence in the spectrogram signaling the presence of laryngealization, f0 contour shows clear f0 drop over the second portion of the vowel. In [ʃarh], on the other hand, the spectrogram shows a clear case of breathiness in the second portion of the vowel, as reflected in the decreased overall acoustic intensity and aperiodic energy at higher frequencies. The same portion is also characterized by the lowering of fundamental frequency in the f0 contour. The modal token [ʃar] shows no evidence of aperiodicity, intensity drop or vowel excursions throughout the vowel. It is also apparent from the spectrogram that the duration of the vowel [æ] in [ʃarʔ] and [ʃarh] is considerably greater than their modal counterpart in [ʃar]. Spectral slope differences are illustrated in Fig. 2 which shows FFT spectra for the vowel [a] in [ʃarʔ] (a), [ʃarh] (b) and [ʃar] (c) sampled at about 45 ms to the vowel offset. The FFT spectrum for [æ] in [ʃarʔ] displays a steeply positive spectral tilt due to a fall off in energy at H1, while the spectrum for [æ] in [ʃarh] shows a most deeply negative spectral tilt due to a considerable loss of energy at H2 and higher frequencies. The modal [æ] in [ʃar] occupies the intermediate ground with its higher frequencies having slightly less amplitude than the fundamental.

## Experiment

### Subjects and tokens

Two sets of monosyllabic words were selected. Each set contained ten words. The words in the first set were of the CVC type, and the words in the second set were of the form CVGC. Five words of the CVGC data contained the glottal fricative /h/, and the other five contained the glottal stop /ʔ/. The words were embedded in ten sentence frames with the main stress falling on the embedded words. 15 native speakers with no known knowledge of linguistics utter colloquially the randomized sentences.

The last 80 ms portion of vowels in each related CVC / CVGC tokens was designated for comparison. The portion of each vowel was tagged at 15 ms time intervals. A Fast Fourier transform (FFT) was calculated over a 25.6 ms window centred at each tag. Values of f0, spectral parameters H1-H2, H1-F2, and vowel duration were computed for vowels in each related CVC / CVGC tokens.



Table 1: means of overall vowel length as well as means of f0 and spectral tilt values at 50 and 20 time intervals for /V/ in CVC and CVGC tokens.

		Mean				Std.dev				t		p	
D(H1-H2)	VCV	50	20	50	20	50	20	50	20				
	V?V	2.4	2.5	0.7	0.7	9.7	12.4	.00	.00				
	V?V	1.3	1.1	0.7	0.5								
D(H1-F1)	VCV	-4.2	-4.1	1.0	0.9								
	V?V	-7.4	-7.1	1.1	1.1	16.5	15.6	.00	.00				
	V?V												
F0	VCV	133	132	10.2	10.5	4.4	4						
	V?V	126	123	10.6	11.8			.00	.00				
	V?V												
L	VCV	96		6.6									
	V?V	141		5.8		-49		m					
	V?V												

		Mean				Std.dev				t		p	
D(H1-H2)	VCV	50	20	50	20	50	20	50	20				
	VhV	2.4	2.1	0.7	0.8	-29	-28	.00	.00				
	VhV	6.2	6.3	0.8	0.8								
D(H1-F1)	VCV	-3.7	-4.1	1.2	1.0								
	VhV	9.6	9.8	1.1	1.0	-75	-75	.00	.00				
	VhV												
F0	VCV	129	129	11.5	9.8	5.7	7.1						
	VhV	121	121	10.3	8.9			.00	.00				
	VhV												
L	VCV	92		8.3									
	VhV	144		5.8		-49		m					
	VhV												

## Recordings

Digitization was performed at 20 KHz. Separate recordings were made for each of the glottal consonants /ʔ/ and /h/. The data recorded for each included 150 tokens (15 speakers  $\times$  2 syllable templates  $\times$  5 words).

## Results and discussion

For each of the parameter H1-H2, H1-F2 and f0, a paired-samples T-test was conducted on the mean values calculated at each time interval across the 80 ms vowel portion. Results indicate that the means of H1-H2, H1-F2, and f0 for the vowels in CVhC and CV?C are significantly different from the vowels in CVC throughout the vowel duration except for the first two time intervals. In addition, T-test shows significant duration differences for modal versus breathy or laryngealized vowel contrast.

A significant finding for CL data acoustic structure was substantial changes in spectral tilt and f0 values during the second half portion of the vowel, as opposed to CVC data. These values are phonetically significant acoustic cues that can be used to recognize words involving CVGC or CVCG sequences. Hence, the result of this study can be used to improve ASR.

## References

- Bijankhan, M. 2000. Farsi vowel compensatory lengthening: an experimental approach. Proc. 5th ICSLP Beijing.
- Blankenship, B. 2002. The timing of non-modal phonation in vowels. *Journal of Phonetics* 30, 163-191.
- Gerfen, C., Baker, K. 2005. The production and perception of laryngealized vowels in Coatzacoaspan Mixtec. *Journal of Phonetics* 33, 311-334.
- Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383-406.
- Shademan, Sh. 2003. Glottal-deletion and compensatory lengthening in Farsi. *UCLA Working Papers in Phonetics*, No, 104, 61-81.

# **Faster time-aligned phonetic transcriptions through partial automation**

Ben Serridge and Luciana Castro

Laboratory of Acoustic Phonetics, Universidade Federal do Rio de Janeiro, Brazil

## **Abstract**

A semi-automatic process for generating time-aligned transcriptions of speech data at the word and phone level is described. At each stage in the process, segment durations are estimated to generate approximate boundary markers, which are then corrected by hand. Corrections at one level are taken into account in the generation of boundaries for the next level, such that the error is reduced at each successive stage. A test implementation based on Praat was applied to a corpus of Brazilian Portuguese and a comparison against a fully manual process revealed a reduction of 54% in the time required to generate phonetic transcriptions and an average error of 21 ms in the time-alignment of phonetic boundaries.

Key words: Brazilian Portuguese, Praat, phonetic transcription, automation tools

## **Introduction**

Linguistics research often relies on access to speech data that has been annotated with time-aligned orthographic and/or phonetic labels. Such corpora are available for heavily studied languages such as English and French, but for most of the world's languages – including major languages such as Brazilian Portuguese – there is very little data available and linguists generally record, transcribe and label their own data as part of their research. The manual transcription process imposes a severe restriction on the amount of data used in the study, and researchers are often forced by time and budget constraints to compromise the robustness of their results.

The ideal solution to the problem is to use an automatic speech recognition (ASR) system configured for forced-path alignment to generate a time-aligned phonetic transcription given the (non-time-aligned) orthographic transcription and a set of grapheme-to-phone rules. Unfortunately, however, there are many languages for which no such ASR system exists, and even for supported languages, the cost, time and technical expertise required to install, configure and successfully apply existing frameworks is prohibitive for many linguistics researchers.

This paper presents a relatively simple process for reducing the time required to transcribe speech data. At each stage in the process, segment durations are estimated using relatively naïve heuristics to generate approximate boundary markers, which are then corrected by hand. Corrections at one level are taken into account in the generation of

boundaries for the next level, such that the error is reduced at each successive stage.

### Transcription framework

The framework described in this paper is based on Praat (Boersma 2008), a commonly available speech analysis framework, and leverages the concept of a transcription tier, which enables several layers of time-aligned linguistic annotations for a single utterance, as shown in Figure 1.

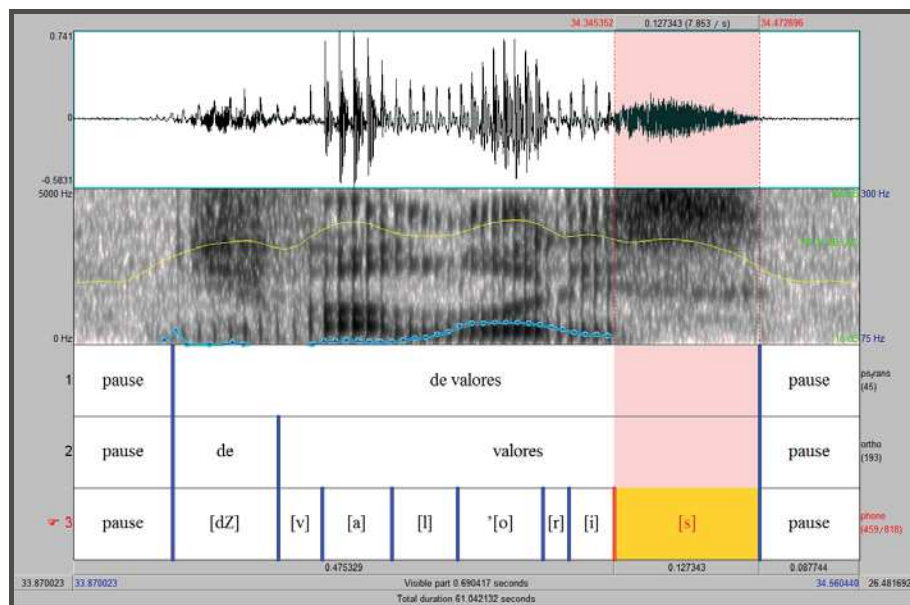


Figure 1. The three transcriptions tiers – phonetic sequence, word-level, and segment-level time-aligned transcriptions – as shown in Praat.

The starting point is a text file in which each line in the file represents the orthographic transcription of a phonetic sequence and line-breaks represent pauses. Taking this file as input, together with the underlying wave file, a Praat script generates the orthographic transcription tier, consisting of alternating intervals representing phonetic sequences (labelled with the transcription itself) and pauses. No pause detection is actually performed on the underlying audio signal; rather, the duration of each pause is fixed (for example, at 500 ms) and the duration of each phonetic sequence is proportional to the number of letters in the orthographic transcription of the sequence, constrained by the overall duration of the phonetic sequences.

Once the phonetic sequence tier has been adjusted by hand, a second script applies a similar procedure to generate a word-level transcription tier,

in which the duration of each word is estimated by multiplying the fraction of letters that the word occupies in the phonetic sequence by the duration of the phonetic sequence as a whole, as given by the previous tier.

Note that up to this point the procedure is fairly language independent. The generation of the phonetic labels, however, requires language-specific rules to predict the set of phones associated with a given orthographic transcription. In this study, the grapheme-to-phone rules described by Silva et al. (2006) were implemented through the use of regular expressions, divided into seven stages, as described in Table 1.

Table 1. Stages in the translation of the orthographic transcription of a word to its phonetic representation (SAMPA).

Phase	Description	Example Transformation
Dictionary Lookup	Handle words whose transcription cannot be predicted by rule.	t á x i → [t] '[a] [k] [s] [i]
Stress Prediction Rules	Mark the vowels that carry primary stress.	c e d o → c 'e d o
Canonical Spelling	Replace known multi-letter combinations with equivalent, unambiguous graphemes.	g 'e s s o → j 'e ç o
Context-dependent Rules	Rules in which the context of the letter determines its mapping.	r a p 'a z → [R] a p 'a [s]
Context-Independent Rules	One-to-one mappings of letters to phones.	[R] a p 'a [s] → [R] [a] [p] '[a] [s]
Standard Phonological Rules	In contrast to the rules applied so far, these operate not on letters but on phones.	[k] '[a] [n] [t] [a] → [k] '[a~] [t] [a]
Regional Phonological Rules	These rules may be applied or not depending on the regional accent of the speaker.	[R] [a] [p] '[a] [s] → [R] [a] [p] '[ai] [S]

The duration of each phonetic unit was calculated based on its average phone duration (Barbosa 1995), scaled appropriately to match the duration of the word as given by the previous tier.

## Results

In order to quantify the efficiency gained by applying the above procedure, two one-minute recordings of Brazilian television news were transcribed using Praat: one recording was transcribed following the procedure described in this paper, and the other without the aid of partial automation.

Table 2. The time required to complete each transcription task, with and without the aid of partial automation, expressed in minutes of transcription time per minute of audio, and the average error (in ms) of the predicted boundaries as compared to the final, hand-adjusted boundaries.

Task	Manual Transcription	Partial Automation	% Reduction	Average Boundary Error (ms)
Alignment of Pauses	18	17	6%	672
Word-Level Alignment	50	27	45%	99
Phonetic Transcription	150	55	63%	21
Overall Transcription Task	218	100	54%	N/A

## Conclusion

The tools and technical know-how required for fully automating linguistic transcription tasks are inaccessible to most linguists and for the vast majority of the world's languages. Partial automation, however, through the procedures described in this paper, can reduce overall transcription time by half, allowing linguists to work with larger corpora or to spend more of their time on analysis and less on the manual tasks involved in transcription.

## References

- Barbosa, P. 1995. Estrutura r tmica da frase revelada por aspectos de produ  o e percep  o de fala. Manuscript of talk given at the XLIII Semin rio do GEL, May 25-27, 1995, S o Paulo, Brazil.
- Boersma, P. and Weenink, D. 2008. Praat: doing phonetics by computer (version 5.0.21, <http://www.praat.org/>).
- Silva, D.C., de Lima, A.A., Maia, R., Braga, D., de Moraes, J.F., de Moraes, J.A., and Resende, F.G.V. 2006. A rule-based grapheme-phone converter and stress determination for Brazilian Portuguese natural language processing. Proc. VIth ITS, September 3-6, 2006, Fortaleza, Brazil.

# **The effects of the acoustic properties of second language vowel production on pronunciation evaluation**

Chris Sheppard

Faculty of Science and Engineering, Waseda University, Japan

## **Abstract**

This study examines the acoustic characteristics which native speakers use as cues when judging speech samples. 17 words produced by 98 (15 native speakers and 83 Japanese learners of English) were collected and evaluated. Following this, the first three formants for each vowel produced were measured and used as independent variables in stepwise multiple regression analyses to explain variance in the evaluations. The results demonstrated that the formants did explain variance in evaluation in many but not all of the words ( $r^2 = 0.0$  to  $.36$ ).

Key words: vowels, pronunciation, evaluation, formants

## **Introduction**

Studies have been using native speaker evaluation of words, sentences and passages to determine factors behind the successful second language pronunciation for more than thirty years (e.g. Oyama, 1976; Piske, 2001; Purcell and Suter, 1980; Sheppard, Hayashi and Ohmori, 2007 just to name a few.) As yet, however, there has been little examination of which acoustic properties of the pronunciation the native speakers were actually evaluating.

Munro and Derwing (1995) have determined that native speaker evaluations were being influenced by suprasegmental over segmental features of learner speech and (Wayland, 1997), analyzing production acoustically, found that the both fundamental frequency and the vowel formants predicted evaluations of English speakers production on non-native Thai words. This study attempts to determine which, if any, of the acoustic qualities of non-native English vowel production by Japanese speakers influence native speaker evaluations

## **Method**

### **Participants**

83 Japanese native speaking volunteers who were studying English were recruited from two universities. Some participants who had attained exceptional levels of pronunciation were invited to participate in the study by the researcher. Another 15 native speakers of American varieties of English took part in the study.

### **Instruments and procedure**

The stimulus words consisted of 26 printed words, 12 pictures, 8 sentences, one paragraph and a question and answer session on a topic selected by the participants. For this study, 17 of the printed words containing vowels (but not diphthongs) were selected for analysis.

The 17 words selected for this study were church, happy, shoes, music, green, insect, fix, goodbye, opinion, nation, open, husband, young, discussion, ask, drink and thoughtful. These represented 10 vowels.

The speech samples were elicited in a studio and recorded at 48,000 kHz using a Sony DAT and a condensing microphone. The samples were presented on a power point slide. After the instructions were given to the participants, they were required to pronounce each word and then click the mouse button to the next slide. There was one practice slide.

### **Analysis**

After uploading and editing the recordings, the 1666 (17 words X 98 participants) samples were presented through Superlab 4.0 to two Native English speakers. Both were teachers with extensive experience with Japanese learners. The pronunciation of each word was rated on a five point scale based on 'how good the pronunciation was'. The Interrater reliability was high relatively at .82.

Next, an acoustic analysis of the vowels was undertaken. This was done by measuring the first three formants of each of the 1666 samples using a Praat script. Next, the formants were entered into a stepwise linear multiple regression analysis with the evaluation of the word as the dependant variable.

### **Results and discussion**

Figure one below describes the mean formants for both Japanese speakers (JS) and Native speakers (NS). Comparing the formants of each word with t-tests, we find that Japanese production differs significantly ( $p < .05$ ) from that of native production in the first formant for church, green, ask and thoughtful, and insect and fix for the second formant. The JS speaker production significantly differs from NS on both formants for the words husband, drink, young, and music.

The results of the multiple regression analyses are summarized on table 1. They demonstrate that the first three vowel formants accounted for between an insignificant amount of variance in evaluation, to a maximum of 36%.

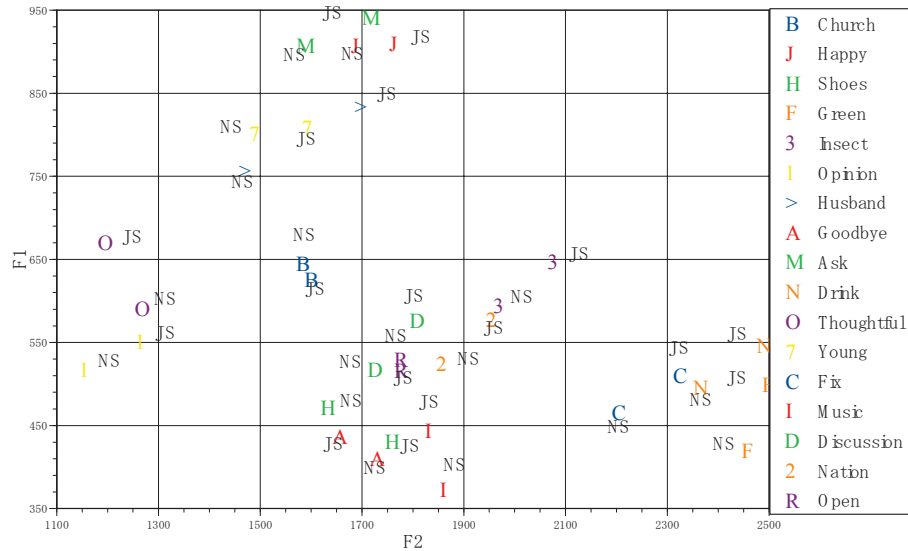


Figure 1. The mean first and second formants of vowels for JS and NS.

Table 1. The results of the step-wise regression analysis for F1, F2 and F3.

Vowel	Item	R <sup>2</sup>		
		F1	F2	F3
i	Green	-	-	-
i	Drink	-	-	-
I	Fix	0.14	0.22	-
e	Insect	0.03	0.25	-
er	Church	0.20	-	-
ae	Happy	-	0.09	0.04
ae	Ask	-	0.07	-
a	Discussion	0.08	0.17	-
a	Young	-	-	-
a	Husband	0.14	-	-
or	Thoughtful	0.17	-	-
u	Goodbye	-	0.09	0.05
u	Shoes	-	0.23	-
u	Music	0.04	0.13	-
schwa	Opinion	-	0.07	-
schwa	Nation	-	0.07	-
schwa	Open	-	0.08	0.04



Thus the first conclusion we can draw from this study is that Native Speaker use F1 and F2 to evaluate pronunciation of vocabulary. This result replicates that of Wayland (1997) but not of Munro and Derwing (1995) who found that suprasegmentals rather the segmental characteristics were being evaluated. However, this result was not gained as a result of an acoustic analysis. Also, the object of evaluation was lengths of multiple words rather than single words.

A second point which requires explanation is the variation in the contribution F1 and F2 had to the evaluation. Firstly, the contribution of the formants was less in multi-syllable words. This makes sense as the longer words have more objects of evaluation. Secondly, the variability in evaluation could have been due to L1 influence or L2 developmental processes or even due to the phonetic environment of the vowel.

### **Conclusion**

This study has demonstrated that variation in the F1 and F2 of vowels in single words can predict the evaluation of those words. However, there is considerable variation in the evaluations which needs to be explained in further research. In addition, the implications for this and the teaching of pronunciation need to be investigated.

### **Acknowledgements**

This work was funded in part by the grant-in-Aid for Scientific Research (C), "Research on the Acquisition Process of Second Language Pronunciation", no. 16520357 from JSPS

### **References**

- Munro M. J. and Derwing T. M. 1995. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45, 73-97.
- Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5, 261-283.
- Piske, T., MacKay, I. R. A., and Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191-215.
- Purcell, E. T., and Suter, R. W. 1980. Predictors of pronunciation accuracy: A re-examination. *Language Learning* 30, 271-287
- Sheppard, C., Hayashi, C., and Ohmori, A. 2007. Factors Accounting for Attainment in Foreign Language Phonological Competence. *International Congress of Phonetic Sciences 2007*, 1597 – 1600, Saarbrücken, Germany.
- Wayland, R. 1997. NonNative Production of Thai: Acoustic Measurements and Accentedness Ratings. *Applied Linguistics* 18, 345-373.

# Frequency effects in language acquisition: a case study of plural forms in Brazilian Portuguese

Thais Cristófaros Silva<sup>1</sup> and Christina Abreu Gomes<sup>2</sup>

<sup>1</sup>Department of Linguistics, Federal University of Minas Gerais, Brazil

<sup>2</sup>Department of Linguistics, Federal University of Rio de Janeiro, Brazil

## Abstract

This paper discusses the acquisition of plural forms of words that end in a back glide in Brazilian Portuguese and may take two plural forms: either –s (the regular suffix) or the suffix –is. Forty children were presented to picture cards containing words ending in a back glide and were asked to identify what they saw. Children's answers presented the suffix –is for the vast majority of words rather than the regular plural suffix –s. Our results indicate that children abstract schemas rather than acquire default rules. The –is suffix has a very high frequency rate for words that end in a back glide which yields to the establishment of robust schemas. We argue that the plural suffix –is is adopted by children due to frequency effects which foments a very robust schema and enhances productivity.

Key-words: acquisition; plural, language use, frequency, schemas

## Morphological storage and language acquisition

The study of the acquisition of regular and irregular inflection is an important issue in the debate as to how morphology is represented in the speaker's grammar/mind. More specifically, this debate addresses how children cope with the competing regular and irregular morphemes during the acquisition process. However, this last issue is interpreted according to the representational model adopted. This paper intends to be a contribution to such a debate by evaluating the acquisition of plural morphemes in Brazilian Portuguese.

In relation to representational models or models of morphological storage, two major trends can be identified: the dual and the single models. The differences between these models reflect different conceptions of how the speakers' grammars are organized. In the dual-processing model regular and irregular inflections differ structurally in terms of storage and processing (Marcus et. al., 1992, Prasada and Pinker, 1993, among others). Regular inflections are derived from a symbolic default rule while irregular inflected words are stored in the lexicon. Thus there is a structural difference between the storage and processing of regular and irregular morphological patterns.

In single-mechanism models, storage and processing of morphological forms are related to a single associative mechanism and regular or irregular forms are treated in the same way as separate entries stored in the lexicon.

Thus, there is no difference in storage and processing of regular and irregular morphological patterns as stated in dual-processing model.

In connectionist models (Macwhinney and Leinbach, 1991) and the network model (Bybee, 1995), all types of morphological properties of words, paradigms and morphological patterns emerge from associations made among the words related in lexical representation. This model suggests that the lexicon is conceived as an organized net of lexical relations which interact in a network fashion. The regularities and similarities observed in linguistic items are properties that contribute to the structure of storage. This assumption is sustained by researches on lexical access. According to Bybee (1995), identity relations are established by phonetic and semantic similarities. Activation of one item spreads to other items. When words are related by parallel semantic and phonological connections, the resulting relations are morphological. In the single-mechanism model type and token frequency play a role in determining representation and type frequency is important in determining the regularity and productivity (Bybee, 1995). Overregularizations found in language acquisition process are interpreted as the overapplication of the most frequent pattern in the lexicon (Bybee, 1995).

This paper intends to show that the network model, which argues that frequency effects play a major role in the organization of Grammar, offers a general and elegant account for the acquisition of plural forms of words that end in a back glide in Brazilian Portuguese (henceforth BP).

### **Inflection of words ending in a back glide in BP**

Words ending in a back glide may take two plural forms in BP: either *–s* (i.e., the regular plural suffix) or the suffix *–is* (which involves a vocalized lateral). Words which take the regular plural suffix *–s* present a back glide in any variety of Portuguese. For example, the singular form *pau* [paw] ‘wood’ has its plural form as [paws] ‘wood(s)’. Words which take the plural suffix *–is* present systematically a back glide in the singular forms for most varieties of BP<sup>1</sup>. For example, the singular for *sal* [saw] ‘salt’ has its plural form as [sais] ‘salt(s)’. The back glide in this later case, i.e., *sal* [saw] ‘salt’, involves lateral vocalization which is already a completed change in the analyzed varieties of BP (Quednau, 1994). Lateral vocalization created then analogous segmental sequences in the end of words: a vowel followed by a back glide (*pau* [paw] ‘wood’ and *sal* [saw] ‘salt’). The plural forms of words containing these word-final analogous vowel-glide segmental sequences are different: either an *–s* suffix, as for *paus* [paws] ‘woods’, or an *–is* suffix, as in *sais* [sais] ‘salts’, occurs. In this paper we will show that the choice between these plural suffixes during language acquisition follows from frequency generalizations over the lexicon.

## Results and discussion

Data were obtained in a controlled experiment with cards presented to the children who were requested to identify the pictures. A card containing a single object was presented to the child and the interviewer pronounced the word in its singular form. Then, immediately after another card was presented to the child with the same object illustrated more than once so that the child was expected to say the plural form. The data collection was recorded using a digital tape recorder. Only the answers with a plural morpheme were considered. The test consisted of 12 real words and 3 nonce words (total 15 words). The three nonce words were presented to participants in the singular form as being an imaginary toy or cartoon character. Token frequency and children's vocabulary were considered when selecting the words to be examined. Frequency count was obtained from a 130 million words corpus and grouped as high frequency and low frequency ([www.projetoaspa.org](http://www.projetoaspa.org)). The 12 real words were grouped as either frequent or infrequent according to the corpus consulted. The audio containing the children's responses was verified by at least two researchers and submitted to statistical analysis of Minitab Program 13 for windows.

We wanted to test at what extent the default regular plural suffix *-s* would occur in words that ended in a back glide. If a default rule was acquired, as suggested by the dual mechanism model, one would expect that children would use preferably the suffix *-s* for words ending in a back glide. This is because the *-s* suffix is the regular plural marker for words ending in vowels and diphthongs in BP. However, this was not the case. In fact children presented the suffix *-is* for the vast majority of words which ended in a back glide (79%). In words where the *-is* suffix was expected, as in *sais* [sais] 'salts', a rate of 94,86% was attested. In words where the *-s* suffix was expected, as in *degraus* [degraws] 'steps', we rather found the rate of 73,51% for the suffix *-is*. Thus, the *-is* irregular plural suffix occurred at a high rate (73,51%) where the regular suffix *-s* was expected. We suggest that these results follow from frequency effects. Plural forms of words that end in a back glide count over a million tokens and 877 types for the *-is* suffix. Concerning the *-s* suffix the token count is just nearly 34.000 words and types count simply 33 words. This shows that the *-is* suffix occurs at a higher rate for token and type frequency when compared to the *-s* suffix. Pseudowords also took the irregular suffix *-is* at a higher rate than the regular suffix *-s*. This indicates that the irregular suffix *-is* is preferred whether or not the word is known.

We suggest that our results indicate that children abstract schemas rather than acquire default rules, thus offering evidence for a network model (Bybee 1995). We argue that the plural marker *-is* is adopted by children due to its type frequency, an effect that foments a very robust schema and

enhances productivity. The –is suffix has a very high type frequency rate for words that end in a back glide which yields to the establishment of robust schemas. That is why the suffix –is is preferred by children. Our results support the view that frequency effect shapes representation and also contributes towards the debate on the acquisition of irregular morphology (Marcus et al, 1992).

### Notes

1. An alveolar lateral occurs in other varieties of Portuguese where the glide occurs in BP : [saw] alternates with [sal] ‘salt’, having both the plural: [sais].

### Acknowledgements

CNPq/Brasil PQ grants 30.33.97/2005-5 and 304056/2007-3.

### References

- Bybee, Joan L. 1988. Morphology as lexical organization. In M. Hammond and M. Noonan (eds.) *Theoretical morphology*. Academic Press. 119-141.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Cognitive Processes* 10. 425-455.
- MacWhinney, B. and Leinbach, J. 1991. Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, 29, 121.
- Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J. and Xu, F. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57 (4).
- Prasada, S and Pinker, S. 1993. Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1-56.

# **Factors influencing perceptual attainment of Japanese geminate consonants by Korean learners of Japanese**

Mee Sonu

Graduate School of Global Information and Telecommunication Studies, Waseda University, Japan

## **Abstract**

Two experiments were carried out to determine the perceptual mechanism which distinguishes geminate consonants of Japanese by Korean learners of Japanese. The first investigated the categorical perception of Japanese geminate consonants. In this experiment, we carried out a listening test to measure perceptual characteristics by Korean learners. The second was an interview of Korean learners on their learning strategies. The experimental results showed that some Korean learners concentrated on the overall rhythmic organization of geminate consonants, the same strategy as Japanese native speakers which had been denied in Min (1987, 2007).

Key words: geminate consonants, perception, learning strategies, Korean learners of Japanese

## **Introduction**

This study investigated the perceptual mechanism of geminate consonants in the second language of Korean learners of Japanese. A large number of previous studies have investigated the perceptual mechanism of geminate consonants by Korean learners of Japanese (Min, 1987, 2007). These studies analyzed the perceptual distinction between geminate consonants and singleton stops. However, all Korean learners do not seem to use the same perceptual cues to distinguish geminate consonants. Accordingly, there is a necessity to determine what these cues are. Namely, we need to analyze the influence of acoustical features on perception. There is an additional necessity to determine the learners' psychological approaches to geminate stop discrimination. We carried out two types of experiments to get answers to the following questions.

First, what are the differences between Korean learners of Japanese and Japanese native speakers? In particular, are there any learners who utilize the same acoustic features to distinguish judgements as native speakers?

Second, what speech features do individual learners use to distinguish geminate consonants from singleton stops? In particular, do good learners of Japanese use the same features as native speakers?

## **Experiment for categorical perception characteristics**

### **Participants**

Eighteen native speakers of Japanese from Tokyo and forty-five Korean Japanese Learners participated in the study.

### **Stimuli and procedure of a perceptual test**

The materials consisted of three pairs of 2 mora and 3 mora words which contrasted singleton and geminate stops (/aka/-/akka/,/saka/-/sakka/,/raka/-/rakka). Each item was produced both in isolation and embedded in a carrier sentence: for example, “*watasiwa \_\_\_\_ to iimasita*” (I said \_\_\_\_). A female Japanese narrator uttered the sentences with an LH pitch accent at a normal speaking rate.

A stimulus set for perceptual experiments was created from the materials by acoustically modifying the duration of the closure between the first and second mora of each word. The length of the stop was modified to provide samples varying by 20ms in length by removing part of the closure. Each subject heard each stimulus a single time and was asked to judge whether he or she heard had heard a 3-mora or 2-mora word.

### **Analysis**

Categorical perception characteristics were represented by two measures. The first one was the boundary point of geminate and singleton stop perception. This was defined as the mid-point in milliseconds between the perception of singleton and geminate consonants. The other was the width of boundary interval where geminate and singleton stop perception co-occurred. As shown in Table 1, boundary points and boundary widths of the Korean learners were compared with Japanese participants.

### **Results and discussion**

Table1 shows the results that both the boundary point and the boundary width were different between the Japanese native speakers and Korean learners on all test pairs. However, there are five Korean learners of Japanese who showed the same boundary point and the boundary width as Japanese native speakers. These results do not coincide with previous experimental results given by Min (1987, 2007).

## **Experiment for learning strategies**

### **Participants**

Forty-five Korean Japanese Learners participated in the study.

### **Stimuli and procedure for understanding learning strategies**

One of the main aims of the experiment is to understand how to acquire Japanese geminate consonants by Korean learners from the psychological viewpoint (Oxford, R.1990). Stimulated recall methodology can be used to prompt participants to recall thoughts they had while performing a task or participating in an event (O'Malley et al.1990).Accordingly, stimulated recall could measure cognitive structure and mental representations and thus could be used for the understanding of learner's strategies in perception.

### **Analysis**

The stimulated recall interviews were transcribed and analyzed qualitatively for commonalities in judgment strategies. Any commonalities were classified with reference to the KJ method developed by a Japanese ethnologist, Jiro Kawakita. The method of KJ utilized utterance cluster logically. The Procedure is as follows. First, the participant's interviews are recorded. Then, the utterances are clustered into related categories. Last, each utterance is placed into groups until there are none left over.

### **Result and discussion**

There were five Korean learners who use the same perceptual characteristics as native speakers among the forty-five participants. The interview results revealed that one of these learners distinguished geminate consonants and singleton stop by perceiving rhythmic differences. This learner took more time to determine the nature of the consonants than the other participants.

On the other hand, the other Korean learners were conscious of the acoustical closure duration of geminate consonants only. These learners were unsure of their judgements and seemed to think that the accurate distinction between geminate consonant and singleton consonant stops was impossible.

These experimental results suggest that, it is important to inform Korean learners that geminate consonants decisions are carried out not merely by absolute closure duration of the geminate consonants but also by Japanese timing perception.



Table 1. Perceptual categorization by Korean learners of Japanese and Japanese native speaker ( $\mu$ : mean,  $\sigma$ : standard deviation, N: number of listeners)

Test pairs	Korean learners of Japanese (N=45)				Japanese native speaker (N=18)			
	Boundary Point		Boundary Width		Boundary Point		Boundary Width	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
/aka/-/akka/	112.5	11.5	42.7	19.4	94.0	14.0	29.5	13.5
/saka/-/sakka/	92.8	11.8	40.5	15.2	82.3	12.2	28.3	9.5
/raka/-/rakka/	93.2	13.3	40.9	19.6	77.8	13.4	29.4	13.6

### Summary

This study shows that all Korean learners of Japanese do not use the same perceptual cues as Japanese Native speakers. However, some Korean learners of Japanese showed the same perceptual characteristics as native speakers. Those learners tend to be conscious of not only closure duration of the geminate consonants but also Japanese timing perception.

### Acknowledgements

This work was supported in part by Waseda University RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for and Scientific Research B, No. 20300069 of JSPS.

### References

- Kawakita, Jiro 1996. Keijeiho o shite katarashimeru. Tokyo, Chuokoronsha Press.
- Min, K. 1987. Perception of the geminate consonants of Japanese by Korean learners. Japanese Language Education, Vol. 62, pp. 179-193. (Japanese)
- Min, K. 2007. The Cause of the Occurrence of Geminate Insertion: Evidence from Korean Learner's Production of the Japanese Voiceless Stops as Geminate. Journal of the phonetic society of Japan, Vol. 11 No. 1, pp 58-70. (Japanese).
- O'Malley, J and A. U. Chamot 1990. Language strategies in second language acquisition. Cambridge, UK, Cambridge University Press.
- Oxford, R. 1990. Language learning strategies: What every teacher should know. New York, Newbury House.

# **Receptive and productive skills of English /l/ and /r/ by Japanese college students in relation to their motivation**

Yuichi Todaka

Department of Intercultural Studies, Miyazaki Municipal University, Japan

## **Abstract**

The objectives of the present study were threefold: (1) the effects of training in /r/-/l/ perceptual identification on /r/-/l/ production; (2) the effects of training in /r/-/l/ production on /r/-/l/ perceptual identification; and (3) the effects of perceptual and production training on learners' motivation to study English. Results indicate that both the perceptual and the production training sessions resulted in improving not only the intended skill but both the receptive and productive skills. Thus, both receptive or productive skill sessions can be used to promote Japanese college EFL learners' English oral communication skills. Nevertheless, the motivation questionnaire shows (1) the importance for teachers to encourage their students to self-study for improving receptive skill and (2) the significance for teachers to understand the efficacy of output practice in small groups to help their students gain confidence in productive skill.

Key words: receptive, productive, motivation

## **Introduction**

Nonnative speakers have extreme difficulty receiving and producing certain non-native phonetic contrasts (e.g., Flege, 1988; Goto, 1971). For instance, Japanese learners of English have difficulty discriminating /r/ and /l/ sounds. According to Ladefoged (2001), /l/ is lateral approximant and /r/ is central approximant while Japanese one is an apico-alveolar tap which is phonetically more similar to flapped /t/ and /d/ in American English (Price, 1981; Vance, 1987 cited in Aoyama et al., 2004). These three phonetic contrasts are, therefore, completely different from one another from both perceptual and articulatory points of view. Nonetheless, the vast majority of Japanese adult second language learners of English recognize both /r/ and /l/ sounds the same as Japanese apico-alveolar tap.

The relationship between speech perception and speech production has also been a long-standing issue in speech science and experimental phonetics (Bradlow, et al. 1997). For example, a rounded vowel in French (/y/) is mispronounced as a high-front vowel (/i/) by Portuguese, but it is mispronounced as a high-back vowel (/u/) by English speakers.

Skehan (1989) suggested that learners' language aptitudes vary individually, and such variation has considerable significance for language learning success. However, assuming that aptitude was innate and not

changeable, other factors, such as attitudes towards the target language and the motivation to learn it, can become the key for L2 acquisition (Baba 1997).

In the present study, we, therefore, investigated: (1) the effects of training in /r/-/l/ perceptual identification on /r/-/l/ production; (2) the effects of training in /r/-/l/ production on /r/-/l/ perceptual identification; and (3) the effects of perceptual and production training on learners' motivation to study English.

## **Method**

### **Subjects study period**

Thirteen female students at Miyazaki municipal university participated in the present study during the spring semester in 2007. Six joined in perceptual training sessions, and the others participated in production training sessions.

### **Assessment**

#### **Assessment for perceptual identification**

The subjects' improvements in perceptual identification ability were assessed by an identification test of a synthetic /rait-/lait/ continuum which was downloaded from the *Advanced Telecommunications Research Institute International's* website (ATRII 2008). In the identification test, the subjects were asked to identify the initial consonant in word which was synthesized artificially, and to make a forced choice between /r/, /l/, and /w/.

#### **Assessment for production**

Auditory judgments by six native English speakers were used to assess the subjects' improvements in their pronunciations of /r/ and /l/ in words.

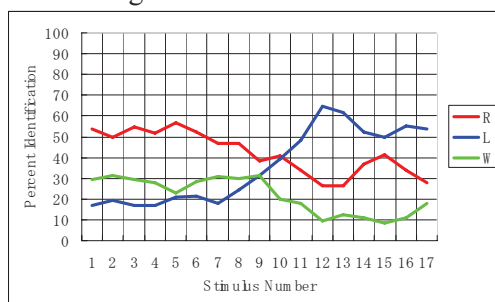
## **Results**

### **Test results**

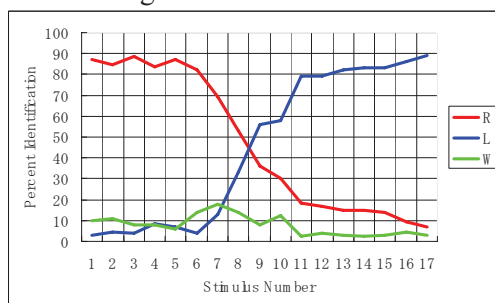
The subjects showed more than 80% correct responses for the stimuli on either side at the post-test (less than 60 % and 70 % for /r/ and /l/ at the pre-test). Rate of responses to /w/ for all stimuli also decreased. The following figure indicates an example of the means of results of identification tests by both groups.

Figure 1. mean results of identification tests and auditory judgments by both groups.

#### Pre-training results



#### Post-training results



Therefore the subjects in both groups showed improvements in their receptive skills

Based upon the native-speaker assessment, the subjects in both groups showed some improvements regardless of types of training they had received; however, we could say that the perceptual group had performed better if you took the deterioration rate into consideration.

#### Questionnaire results

We gave the subjects the same five-point scale questionnaire to the subjects twice to investigate any changes in their motivation to study English. Thirteen questions on the questionnaire were taken from Kiss and Nikolov (2006), and were translated into Japanese.

The results indicate an interesting dichotomy between the perceptual and the production group members. Thus, the importance of cooperative learning in constructing necessary L2 skills (Kaufman 2004, Slavin 2003) and the significance of intrinsic motivation (Wu 2003, Noels et al. 2000) have been reconfirmed in this study.

### Conclusion and limitations

Both perceptual and production training sessions resulted in helping the subjects improve both their perceptive and productive skills to some extent. Nevertheless, it is much more important to consider the subjects' motivation to further study their receptive and productive skills of the targeted consonants, as our ultimate goals as instructors is to assist learners in becoming autonomous learners, which, in turn, results in long-term retention of learned materials. If so, individualized teaching method should not be overlooked as was found in the present study. Both cooperative learning and customized learning go hand in hand in helping our learners acquire the target sounds.

### References

- Advanced Telecommunications Research Institute International. 2008. Internet Koukai Jikken. Retrieved from <http://atrcall.isd.atr.co.jp/>
- Aoyama, K., Flege, J. E., Guion, S. G., Yamada, R. A., and Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32 pp.233-250
- Baba, T. (ed.) 1997. *Eigo Speaking Ron*. Tokyo: Kagensya.
- Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R. and Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Acoustical Society of America* 101, 4, 2299-2310.
- Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuro-psychologia* 9, 317-323.
- Flege, J.E. 1988. The production and perception of foreign language speech sounds. In Winitz, H. (ed.) *Human Communication and Its Disorders: A Review*, 233-271. Norwood, NJ: Ablex.
- Strange, W. (ed.) *Language Research*. Maryland: York Press.
- Kaufman, D. 2004. Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics* 24, 303-319.
- Kiss, C. and Nikolov, M. 2006. Developing, piloting, and validating an instrument to measure young learner's aptitude. *Language Learning* 55 1, 99-150.
- Noeuls, K., Pelletier, L., Clement, R. and Vallerand, R. 2000. Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning* 50, 57-85.
- Skehan, P. 1989. *Individual Differences in Second Language Learning*. London: Edward Arnold.
- Wu, W. 2003. Intrinsic motivation and young language learners: The impact of the classroom environment. *System*, 31, 501-517.

# Objective evaluation of second language learner's translation proficiency using statistical translation measures

Hajime Tsubaki<sup>1</sup>, Keiji Yasuda<sup>2</sup>, Hirofumi Yamamoto<sup>2,3</sup> and Yoshinori Sagisaka<sup>1,2</sup>

<sup>1</sup>GITI / Language and Speech Science Research Laboratories, Waseda University

<sup>2</sup>NiCT/ATR, Japan

<sup>3</sup>School of Science and Engineering, Kinki University, Japan

## Abstract

For objective evaluation of the second language learner's translation proficiency, we tested two measures commonly used in statistical machine translation. One is word n-gram probability of a target language to measure likelihood of translated sentences as a target language. The other one is translation probabilities from source language sentences to target language sentences to measure translation accuracies between these sentences. The subjective proficiency scores of Japanese learners were compared with these objective measures extracted from English sentences translated by them. Statistical analysis showed no correlation of word n-gram probabilities but high correlation of translation probabilities which suggests the usefulness for objective evaluation of learner's translation proficiency.

Key words: objective proficiency evaluation, word n-gram, translation probability

## Introduction

In language processing, quite a few studies have been carried out to automatically evaluate machine translation. Statistical measures such as NIST and BLEU have been extensively studied to emulate human's evaluation characteristics in comparison between man and machine (K. Papineni et al., 2002). These statistical objective measurements have also been effectively used in automatic evaluation of human learner's translation capabilities. In the evaluation of second language proficiency, these measures have been employed to evaluate proficiency by computing statistical differences between learner's sentences and native's ones by replacing an output from a translation system (Yasuda et al., 2003).

Though these studies have shown possibilities of objective evaluation for second language proficiency using comparative measures, they require correct answer sentences of a test set. It is quite laborious to prepare correct answers for every test sentences. To be free from this tedious data collection, we tried to use two measures, word n-gram probability and translation probability, employed in machine translation for learner's proficiency evaluation. If some of statistical translation measures are useful in the

evaluation of learner's sentences, we need not be bothered to prepare answer sentences for every test sets.

### Measures for object evaluation

To estimate second language learner's proficiency from translated sentences, we need measures used by native raters. Considering their proficiency rating, we can find that they use multiple criteria such as (1) Word correspondences in translation, (2) Likelihood as an English expression, (3) Grammaticality, (4) Recoverability or seriousness of miss-translation and (5) Adequacy of corresponding target word selection. For objective evaluation, it is ideal to define a quantitative measure integrating all these criteria. However, in reality, it is not so easy to quantify what factors are relating how in subjective evaluation. It is difficult not only to list up all factors but also to prepare reasonable amount of learner's corpora to get reliable results. In this study, as a first step, we expect that two measures used in statistical translation, word n-gram probability and translation probability, can reflect the first two criteria.

As well known, in statistical machine translation from a Japanese sentence  $j$  to an English sentence  $e$ , English sentence  $e$  that maximizes  $P(e|j)$  in all translation candidates is selected by using two statistics  $P(e)$  and  $P(j|e)$  as expressed in the following equation.

$$e = \underset{\text{all candidate}}{\operatorname{argmax}} \quad P(e|j) = \underset{\text{all candidate}}{\operatorname{argmax}} \quad P(j|e) \cdot P(e)$$

where  $P(j|e)$  stands for translation probability and  $P(e)$  corresponds to occurrence probability of English sentence  $e$ . Two measures that we use for our analysis correspond to these two probabilities.

The translation probability  $P(j|e)$  is calculated using IBM Model 1, word-based translation model (Peter E Brown et al., 1993).  $P(j|e)$  is obtained by word-to-word translation probabilities between the English and Japanese sentence. On the other hand, English sentence probability  $P(e)$  is approximated by word 3-grams probabilities  $P(w_i|w_{i-2}, w_{i-1})$  as follows.

$$P(e) \approx \prod P(w_i|w_{i-2}, w_{i-1})$$

In the evaluation, we apply the above calculation formula by considering an English sentence translated by a learner as a translation candidate in statistical translation.

### Evaluation experiment using two statistical measures

We calculated English word n-gram probability and translation probability from a Japanese sentence to an English one to evaluate their effectiveness for sentence accuracy and learner's proficiency. Using these probabilities, we got correlation scores between these measures and subjective scores for test

set sentences. Finally, we calculated the correlation between learner's proficiencies and objective scores using an effective probability.

### Experimental setup

For the evaluation experiment, we employed a sentence set consisting of Basic Traveler's Expression Corpus (BTEC) (Takezawa et al., 2002) and learner's corpus. Translation probability  $P(j|e)$  and word 3-grams probability  $P(w_i|w_{i-2}, w_{i-1})$  were calculated using the BTEC. The learner's corpus consists of 473 source Japanese sentences translated by 21 learners with different English proficiencies and evaluated in five scales by a Japanese-English bilingual rater based on evaluation criteria (S:Native, A:Good, B:Fair, C:Acceptable and D:Nonsense).

### Experimental results and discussions

Figure 1 shows averages and standard deviations of (a) translation probabilities and (b) word n-gram probabilities over all sentences belonging to each subjective scoring category from D (lowest) to S (highest). As Figure 1 shows, the translation probability average increases as subjective score becomes high. On the other hand, the word n-gram probabilities show no correlation between subjective scores. The positive correlation in translation probability indicates its usefulness in the objective evaluation of sentences.

To confirm the validity of translation probability for the objective evaluation of learner's proficiency, we calculated correlation between the averages of translation probabilities for each learner and the learner's TOEIC scores. The correlation was turned out to be 0.287. By analyzing the data, we found that this low correlation results from lower correlations in short sentences. To quantify the effects of sentence length, we measured the correlations of subgroups divided by their sentence length. As shown in Table 1, we could find the increase of correlations between the averages of translation probabilities and TOEIC scores in proportion to sentence length.

These results suggest that the translation probability is useful for objective evaluation of learner's proficiency. We need further studies for more efficient use of it by taking test sentence length or sentence complexities into account. While, the word n-gram possibility turned out to be of no use even for sentence evaluation. This result is quite different from the expectations from related previous works (Yasuda et al., 2003). As word statistics were differently used in this work, the word n-gram possibility might not show any differences. Moreover, all learners tend to translate using word sequences that they are familiar with, their occurrence possibilities may not vary so much according to their proficiencies. We should employ lexical information such as word difficulty ranking or expressions directly reflecting proficiencies.



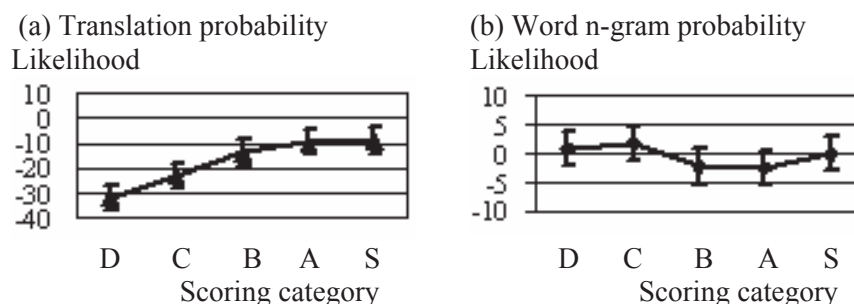


Figure1: Average and standard deviation of translation probabilities and word n-gram probabilities over all sentences belonging to each subjective scoring category.

Table1. Increase of the correlation score between translation probabilities and learner's TOEIC scores to the sentence length in words (Japanese)

Japanese sentence length	1 ~ 5	6 ~ 10	11 ~ 15	16 ~ 20	21 ~
Correlation score	0.240	0.300	0.351	0.404	0.467

## Conclusions

To obtain effective measures for objective evaluation of learner's second language proficiency without being bothered by tedious correct data collection, we have tested the availability of translation probability and word n-gram probability. The analysis experiment showed the usefulness of translation probability for sentence translation evaluation and learner's TOEIC scores. We also found that further specification of measure would increase its effectiveness. We will continue to find parameterizations of other measures that we have not yet used together with effective use of learner's data by themselves.

## References

- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. Proc. of ACL2002, pp. 311-318.
- Keiji Yasuda, Eiichiro Sumita, Seiichi Yamamoto, Masuzo Yanagida, Kikuo Maekawa, and Fumiaki Sugaya. 2003. A Proposal for Automatically Gauging of English Language Proficiency. IPSJ SIG Technical Report, Vol.2003-NL-155: 65-70.
- Peter E Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. Computational Linguistics 19(2):263-311
- Toshiyuki Takezawa, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto. 2002. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. LREC2002, pp.147-152.

# Automatic labeling of prosody

Agnieszka Wagner

Department of Phonetics, Adam Mickiewicz University, Poland

## Abstract

The paper proposes a framework for automatic prosody labeling. The labeling involves detection of the location of accented syllables and phrase boundaries, and recognition of pitch accent and boundary tone types. A number of classification models are designed to perform these tasks on the basis of small vectors of acoustic features. The models achieve high accuracy and their performance is comparable to the results reported in other studies on automatic prosody labeling and to inter-labeler consistency in manual labeling of prosody.

Key words: prosody, description, automatic recognition/detection

## Introduction

In recent years a growing interest in automatic labeling of prosody is observed, which can be attributed to the development of speech applications such as speech synthesis or recognition. Generally, all types of speech applications require speech corpora which have to be provided with appropriate annotation. Manual annotation of prosody is laborious, time-consuming and not entirely consistent e.g., Pitrelli et al. (1994), Grice et al. (1996), Yoon et al. (2004). A solution to these problems is development of models providing the labeling of prosody automatically e.g., Wightman and Ostendorf (1994), Kiessling et al. (1996), Sridhar et al. (2007).

In the next sections methods capable of identifying the prosodic structure of utterances and providing the description of prosodic events (pitch accents, boundary tones) on the surface-phonological level are proposed. They were designed on the basis of a subset of the unit selection speech corpus used in the Polish module for BOSS, Breuer et al. (2000). The speech material consisted of 1052 utterances read by a professional male speaker in a news-reading style, but examples of expressive speech were also provided. The corpus was automatically segmented on the phoneme/syllable/word level; Stress was assigned from rules. Prosodic labeling was done manually.

Decision trees and neural networks (MLP, RBF and linear networks) were applied to solve the detection/recognition problems. In the next sections only the results achieved by the best models are reported. All the models were designed using the *Statistica Neural Network* package available in Statistica 6.0 (2001).

**Automatic labeling of boundary tones**

It involves two steps, namely detection of phrase boundary location and recognition of boundary tone type. The former is performed on the word-level i.e. only word-final syllables are taken into account, which ensures that only one boundary per word can be identified. The latter is performed only for phrase-final syllables. The inventory of boundary tones consists of 2 rising (labeled 2,? and 5,?) and 3 falling boundaries (2,, 5,, and 5,!). They are distinguished on the basis of direction and amplitude of the distinctive pitch movement and scaling of f0 targets at the start and end of the movement. Boundary tones labeled 2,? and 2,, are associated with minor phrase boundaries, whereas 5,, 5,? and 5,! – with major phrase boundaries. The inventory is part of the prosody description on the surface-phonological level which encodes both melodic and functional aspects of prosody. As shown in Wagner (2008) this description provides information which is highly significant to the estimation of pitch target level for F0 generation in speech synthesis.

**Detection of phrase boundary location**

For this task the following acoustic features were used: 1) relative duration of the nucleus of the current and previous syllable, 2) relative duration of the current syllable, 3) F0 features determined for the vocalic nucleus - tilt value, rising amplitude, overall F0 level and slope. The tilt and amplitude parameters were calculated as in Taylor (2000). Relative durations were calculated as in Rapp (1998) and were used to eliminate the effect of syllable structure and/or vowel type on the observed duration.

The highest accuracy of phrase boundary detection was achieved with a radial basis function (RBF) network with 23 neurons in the hidden layer. The network was trained using 5844 syllables and tested on a subset consisting of 1000 syllables. In the test sample the location of phrase boundary was identified with **81,6%** accuracy, whereas syllables of a non-final position in the phrase were correctly identified in **79,3%**.

**Recognition of boundary tone type**

The vector of acoustic features used in this task consisted of: 1) syllable-final F0 value, 2) overall F0 level on the previous syllabic nucleus, 3) direction of the pitch, and 4) distance to the next silent pause measured in the number of syllables.

The RBF network with 54 neurons in the hidden layer had the best performance – it achieved an overall accuracy of **87,6%** (test sample). The network was trained on a subset of 1132 syllables and tested on 377 syllables. The boundary tones labeled 5,, were identified with the highest

accuracy i.e. 98,6%, whereas boundaries labeled 2,. were correctly recognized in 70,13%.

### **Automatic labeling of pitch accents**

The first task involves detection of accented syllable position and it is performed on the word-level i.e. only stressed syllables are taken into account. In this way only one accented syllable per word can be identified. Then, the types of pitch accents distinguished on the surface-phonological level are recognized. In this latter task only accented syllables are taken into account. The inventory of pitch accents includes 2 rising accents (labeled L\*H and LH\*), 2 falling accents (H\*L and HL\*) and 1 rising-falling accent (LH\*L). They are distinguished on the basis of the direction of the pitch movement, timing of the F0 peak relative to the accented syllable onset and range of an f0 change on the vocalic nucleus. Together with the inventory of boundary tones, the pitch accent inventory constitute the description of prosody on the surface-phonological level.

### **Detection of accented syllable location**

The detection of accented syllable position was based on 2 duration features including relative duration of the syllabic nucleus and syllable, and 3 F0 features describing the amount of pitch variation on the syllable, peak height and tilt value of the syllable. Altogether 6417 stressed syllables were used in the experiments with 3929 accented syllables among them.

MLP and RBF networks performed significantly better than the decision tree or the linear network. The MLP network achieved **81,65%** accuracy in the detection of accented syllables and **81,79%** in the detection of unaccented syllables (test sample). The results for the RBF network were **82,14%** and **81,76%** respectively.

### **Recognition of pitch accent type**

The recognition of pitch accent types required a larger acoustic feature vector consisting of parameters describing the amplitude of the pitch movement on the vowel, direction of the pitch movement (calculated as a difference in mean F0 on the vowel between the current and next syllable), tilt value determined in a 2-syllable window containing the current and next syllable, F0 peak, minimum and mean associated with the accent and normalized relative to the F0 mean in the phrase.

The best results were achieved with a classification tree including 27 splits and 28 terminal nodes. The tree was designed using QUEST classification programme available in Statistica 6.0. 2754 syllables were used for training and 917 for testing. The tree performed with an overall

accuracy of **81,63%** (test sample). LH\*L accents were recognized with the highest accuracy i.e. 89,3%, whereas LH\* accents were correctly identified in 70,2%.

### Discussion and conclusions

The models presented in this paper recognize the prosodic structure of utterances and provide the surface-phonological description of prosody in terms of different types of pitch accents and boundary tones.

The performance of the models is comparable to the results reported in other studies on automatic prosody labeling e.g. Rapp (1998), Sridhar et al. (2007) and inter-labeler consistency in manual transcription of prosody e.g. Grice et al. (1996). The advantage of the models proposed in this paper is that they require only small vectors of acoustic features (as opposed to e.g. 276 features used in Kiessling et al. (1996)) which can be easily derived from utterance's acoustics.

### References

- Breuer S., Stober K., Wagner P. and Abresch J. 2000. Dokumentation zum Bonn Open Synthesis System BOSS II. Project report, IKP, University of Bonn.
- Grice M., Reyelt M., Benzmueller R. and Mayer J., Batliner A. 1996. Consistency in transcription and labeling of German intonation with GToBI. Proc. of the 4th ICSLP, 1716-1719, Philadelphia, USA.
- Kiessling A., Kompe R., Batliner A., Niemann H., Nöth E. 1996. Classification of Boundaries and Accents in Spontaneous Speech. Verbmobil-Report 156, University of Erlangen-Nuremberg, University of Munich, August 1996.
- Pitrelli J.F., Beckman M.E. and Hirschberg J. 1994. Evaluation of prosody transcription labeling reliability in the ToBI framework. Proc. of the 3rd ICSLP, 123-126, Yokohama, Japan.
- Rapp S. 1998. Automatic Labeling of German Prosody. Proc. of the 5th ICSLP, Sydney, Australia.
- Sridhar V. K. R., Bangalore S., Narayanan S. 2007. Exploiting Acoustic and Syntactic Features for Automatic Prosody Labeling in a Maximum Entropy Framework. IEEE Transactions on Audio, Speech and Language Processing, 16(4), 797-811.
- Statistica 6.0 2001. Statistica for Windows [computer program], StatSoft, Inc., Tulsa
- Taylor P. 2000. Analysis and synthesis of intonation using the Tilt model. J. of Acoust. Soc. Am. 107(3), 1697-1713.
- Wagner A. 2008. A comprehensive model of intonation for application in speech synthesis. PhD thesis, Institute of Linguistics, Adam Mickiewicz University.
- Wightman C. and Ostendorf M. 1994. Automatic labeling of prosodic patterns. IEEE Transactions on Audio, Speech and Language Processing, 2(4), 469-481.
- Yoon T. J., Chavarría S., Cole J. and Hasegawa-Johnson M. 2004. Intertranscriber Reliability of Prosodic Labeling of Telephone Conversation Using ToBI. Proc. of the 8th ICSLP, 2729-2732, Jeju Island, Korea.

# **Name dominance in spoken word recognition is (not) modulated by expectations: evidence from synonyms**

Andrea Weber<sup>1</sup> and Alissa Melinger<sup>2</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics, The Netherlands

<sup>2</sup>School of Psychology, University of Dundee, Scotland, United Kingdom

## **Abstract**

Two German eye-tracking experiments tested whether top-down expectations interact with acoustically-driven word-recognition processes. Competitor objects with two synonymous names were paired with target objects whose names shared word onsets with either the dominant or the non-dominant name of the competitor. Non-dominant names of competitor objects were either introduced before the test session or not. Eye-movements were monitored while participants heard instructions to click on target objects. Results demonstrate dominant and non-dominant competitor names were considered for recognition, regardless of top-down expectations, though dominant names were always activated more strongly.

Key words: spoken-word recognition, synonyms, pre-exposure, eye tracking

## **Introduction**

Eye movements to displayed objects are one of the most informative methods for studying spoken-word recognition as it happens. Eye-movement studies have, for example, confirmed phonological cohort activation (e.g., Allopenna et al., 1998): when hearing the name of an object, for instance *candle*, listeners fixate phonologically similar objects like *candy* more than phonologically dissimilar objects. Cohort activation is furthermore modulated by lexical frequency (e.g., Dahan et al., 2001, Weber and Crocker, 2006): high-frequency cohort competitors are fixated more than low-frequency ones. Semantic information from preceding contexts also constrains cohort activation (e.g., Dahan and Tanenhaus, 2004): listeners no longer fixate competitor pictures when a preceding verb constrains upcoming material. Thus, upon hearing *the woman lights the...*, only suitable objects like candles but not candy will be fixated. The present eye-movement studies investigated the interaction of bottom-up frequency effects with top-down context effects in German.

Two important changes were incorporated compared with previous studies. Rather than varying lexical frequency between different conceptual objects (e.g., comparing high-frequency competitor *bed* with low-frequency competitor *bell*), and thereby introducing the possibility that items vary along dimensions other than frequency, we investigated name dominance effects for single conceptual and pictorial representations. To this end, we

selected competitor objects with two synonymous names that varied in dominance (e.g., ‘accordion’, dominant *Akkordeon* and non-dominant *Ziehharmonika* in German). Some pictorial displays of synonyms can potentially bias towards a particular word form, thereby surpassing lexical frequency effects (e.g., the picture of a magician and a wizard can look differently). We avoided such items, and rather established via a naming pre-test that synonym pictures evoked two names with varying frequency. Secondly, testing the influence of word expectation within a sentence context usually complicates effect localization; we therefore modulated listeners’ expectations with pre-exposure to a particular word form in isolation. In particular, listeners were introduced to the non-dominant names of competitor pictures prior to the eye-tracking session.

If lexical frequency effects operate for synonyms as they do for other words, more looks to the same competitor picture should be observed when its dominant name is phonologically related to the target picture (*Akkordeon* for target *Akrobat*, ‘acrobat’) than when its non-dominant name is (*Ziehharmonika* for target *Zielscheibe*, ‘target disc’). If pre-exposure further modulates listeners’ expectations for non-dominant competitor names, either by inhibiting high-frequency names or by priming low-frequency names, the magnitude of the competitor activation should vary with pre-exposure.

## **Experiment**

### **Method**

#### **Participants**

Sixty-four native speakers of German, all students at Saarland University, took part in the experiment for monetary compensation. Half of them participated in Experiment 1a (target similar to non-dominant competitor name), the other half in Experiment 1b (target similar to dominant competitor name).

#### **Materials**

Twenty competitor pictures with two synonymous names varying in dominance (‘accordion’, *Akkordeon* and *Ziehharmonika* in German) were displayed together with target pictures that were phonologically related to either the non-dominant competitor name (‘target disc’, *Zielscheibe*, Experiment 1a) or the dominant competitor name (‘acrobat’, *Akrobat*, Experiment 1b). Two unrelated distractor pictures were added to each display (e.g., ‘pear’, *Birne*, and ‘church’, *Kirche*, see Figure 1). Forty filler trials similar in setup but with no phonemic overlap between object names were prepared in addition.



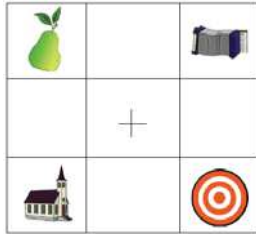


Figure 1. Example display from Experiment 1a.

### Procedure

Instructions to click on objects (e.g., ‘Click on the target disc’) were presented auditorily. Trials were divided into two blocks. For one test block, the complete set of pictures was shown to the participants before the experiment; for the other block, participants were additionally familiarized with the picture names (non-dominant names for competitor objects).

Listeners' eye movements were monitored while they were listening to the instructions. A camera on the participants' dominant eye provided the input to the eye tracker (SMI Eyelink head-mounted). Onset and offset times and spatial coordinates of eye fixations were recorded.

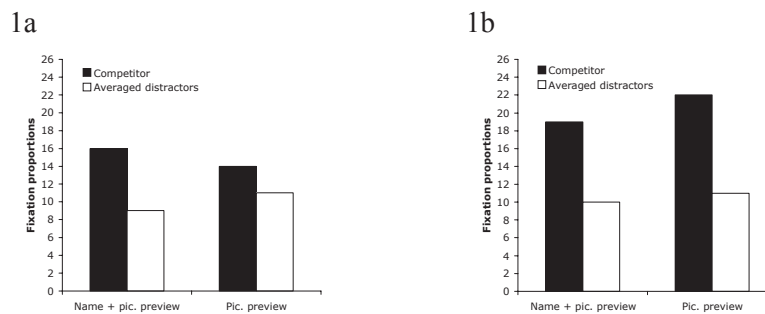


Figure 2. Average fixation proportions for competitor and distractor pictures between 250 and 700 ms after target onset in Experiments 1a and 1b.

### Results

Figure 2 presents the averaged fixation proportions for cohort competitors and distractors between 250 and 700 ms after target word onset in Experiments 1a and b. In this time window competitor activation is commonly observed in eye-tracking studies.

When no name expectation was induced in the picture pre-view, listeners looked at the competitor more often when the target was phonologically similar to the dominant name (Exp. 1b) than the non-dominant name (Exp. 1a;  $t(62) = -2.13$ ,  $p < .04$ ), thus replicating previous frequency effects. However, when we biased listeners to expect the non-dominant name, this



difference disappeared ( $t(62) = -.731, p > .4$ ), suggesting that pre-exposure increased the availability of the non-dominant name. However, the size of the cohort effect was not significantly modulated by expectation. While there was a numerical decrease in the looks to competitors when targets were phonological neighbours of the dominant name, this difference was not statistically reliable, suggesting that pre-exposure does not restrict the candidate set considered by the word recognition system.

### Discussion

Eye movements showed that, indeed, dominant synonym names were activated more strongly than non-dominant synonym names. This finding is in line with previous results for nouns varying in lexical frequency, and generally supports the claim that long-term accrued experience with a word influences the recognition process (e.g., Goldinger, 1998). However, while we did find evidence that pre-exposure increases the availability of the primed name, we did not find evidence that it inhibited the accessibility of the alternative name. This result contrasts recent eye-tracking studies who found an early effect of sentential context on spoken-word recognition. Rather it suggest that lexical access is initially form-based and context-independent (e.g., Norris et al., 2000).

### References

- Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Dahan, D. and Tanenhaus, M. K. 2004. Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Language and Cognitive Processes*, 30, 498–513.
- Dahan, D., Magnuson, J., and Tanenhaus, M. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- Goldinger, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Norris, D., McQueen, J. M., and Cutler, A. 2000. Merging information in speech recognition: feedback is never necessary. *Behavioral & Brain Sciences*, 23, 299–370.
- Weber, A. and Crocker, M. W. 2006. Top-down anticipation versus bottom-up lexical access: Which dominates eye movements in visual scenes. 19th Annual CUNY Sentence Processing Conference, New York.

## Vocal stereotypes

Melanie Weirich

Centre for General Linguistics (ZAS), Germany

### Abstract

German speakers receive different ratings in their perceived vocal attractiveness and personality characteristics. An experiment was conducted to evaluate the perceived personality attributions of 32 German listeners and find acoustic cues that correlate to these attributions. The attributed opposition pairs on which the speakers were judged were reduced by a cluster analysis to two factors: *dominance* (e.g. “confident”, “competent”) and *benevolence* (e.g. “sensitive”, “helpful”). The acoustic investigation of the voices of 4 male speakers revealed the impact of several acoustic parameters on the ratings. Among others, the harmonics-to-noise ratio (HNR) and breathy vocal onsets were most important for high ratings on benevolence. High ratings on dominance showed strong correlations to low mean F0 and strong glottal impulses.

Key words: vocal stereotypes, vocal attractiveness, personality attributions, dominance, benevolence

### Introduction

For more than 50 years, studies on the attribution of personality characteristics to voice quality descriptions have been conducted (Allport and Cantril 1934, Brown 1982, Zuckerman and Driver 1989, Sendlmeier and Siegmund 2005). While these attributions have been relatively consistent between listeners who had to judge different voices, the reason for this consistency has remained somewhat unclear.

Several of these studies found correlations between subjective evaluations of a voice (e.g. *shrillness*, *loudness*, *throatiness*) and perceived vocal attractiveness. However, the correlations to objective acoustical measures were rather low (Zuckerman and Miyake 1993) or contradictory (e.g. influence of F0 excursions on the factor competence by Brown (1982) and G  linas-Chebat (2003)). Judgements of several personality factors were matched with different voice parameters. A high speaking rate for example showed a positive effect on the factors competence, activity and attractiveness (Brown 1982, G  linas-Chebat 2003, Sendlmeier 2005) but a negative effect on benevolence (Sendlmeier 2005).

Thus, the aim of this study is to investigate “vocal stereotypes”, i.e. consistent personality attributions and vocal attractiveness ratings of several speakers by different listeners. In addition to that, we will discuss auditory and acoustic correlates to the subjective judgements of the raters.

## Method

We recorded 25 male native speakers of German reading a passage of “*The little Prince*” by Antoine de Saint-Exupéry. 32 male and female listeners rated each speaker on 15 attribute opposition pairs such as extroverted – introverted, competent – incompetent, honest – dishonest on a scale from 1–7. This method called “semantic differential” extends back to Osgood, Suci and Tannenbaum (1957). In total, 12.000 judgements were elicited (25 speakers x 15 attribute-pairs x 32 listeners).

A cluster analysis was performed on the rating data, reducing the dimensions on which speakers were judged (attribute opposition pairs) to two factors. Orientated on earlier literature we named the factors *dominance* (“confident”, “strong”, “competent”, etc.) and *benevolence* (“sensitive”, “helpful”, “warm”, etc.). This finding is consistent with results presented in recent studies (Sendlmeier and Siegmund 2005, Zuckerman and Driver 1989).

This study extends previous work by acoustically investigating the voices of 4 selected speakers that have received rather unambiguous ratings in order to find objective acoustic cues correlated to the consistent attributions of different listeners: The speakers were either high or low on both factor values, or high on one but low on the other.

Mean F0 and formants were measured but special attention has been given to previously rather neglected acoustic measures such as “harmonics-to-noise ratio” (HNR), “Relative Average Perturbation” (RAP) for jitter, and “Amplitude Perturbation Quality” (APQ) for shimmer. RAP describes micro-fluctuations in the mean F0 and reflects the difference between the calculated mean value of three neighboured oscillations and the actual value. Shimmer describes micro-fluctuations in the sound’s intensity. Other parameters (breathiness, sonority, harshness) were auditory judged based on acoustic evidence such as VOT, periodicity, harmonics, glottal impulses and vocal onsets (VRT).

## Results and discussion

The speakers were judged significantly different in their vocal attractiveness and personality attributions based on their voices. The high interrater reliabilities show the congruency of the ratings: Cronbach-alpha = 0.97 (*dominance*), 0.94 (*benevolence*) and 0.95 (attractive voice).

### The attractive voice

Our results confirm the findings of Zuckermann and Driver (1989) that listeners judge speakers to be different with regard to vocal attractiveness. In addition to that, the data shows, that the correlation between perceived vocal

attractiveness and the attributed personality factors *dominance* and *benevolence* differ in their strength: The association between attractive voice to *dominance* was stronger than to *benevolence*, cp.  $r = 0.82$  vs.  $r = 0.68$  (Pearson correlation,  $p < .01$ , two-sided)

### The personality factors *benevolence* and *dominance*

High jitter values correlate with low ratings of *benevolence*, whereas high shimmer values which are associated with breathiness have a positive effect to *benevolence* scores. Further, soft and breathy vocal onsets and high values for HNR showed a strongly positive effect to the judgement of *benevolence*.

Disturbed periodicity and harshness affect both factors negatively, corresponding to results from Teshigawara (2003) and Laver (1994). The correlation of high ratings on *dominance* and low mean F0 and laryngealisation could also be confirmed (Laver, 1994, Zuckerman and Miyake, 1993). Furthermore, correlations between the attribution of dominance and the existence of numerous harmonics, strong glottal impulses and several intense formants were found. These observations could refer to the connection of spectral tilt of a voice to the attribution of dominance.

Table 1 shows the impact of spectral parameters on the ratings of 4 speakers. Measured mean values from Yumoto, Gould and Baer (1982) for HNR (11.9 dB), from Walton and Orlikoff (1994) for RAP (0.28 %) and from Davis (1979) for APQ5 (5.97 %) can be used as reference.

Table 1. Measurements of Jitter, Shimmer, HNR and mean F0 of 4 speakers scoring high or low on *dominance/benevolence* and a combination thereof.

speaker	dom.	ben.	HNR (dB)	RAP (%)	APQ5 (%)	mean F0 (Hz)
1	high	low	8.30	0.94	4.22	99.28
2	low	low	6.95	1.34	5.24	129.30
3	low	high	10.80	0.37	6.88	133.00
4	high	high	11.60	0.34	5.75	89.40

Several acoustic patterns of the voice could be linked positively or negatively to both factors of personality. Especially features of laryngeal settings and here above all breathiness seem to have the greatest influence to positive attributions in regard to the factor *benevolence*. High scores in the factor *dominance* were more closely related to low mean F0 with laryngeal voice, glottal and pharyngeal laxness and a high sonority reflected through the existence of numerous harmonics and strong glottal impulses. Laryngeal and supra laryngeal settings but also the speaking style played a role.

## Conclusion

The study confirmed the existence of vocal stereotypes for male speakers of German and explored various spectral parameters associated with the raters' conforming judgements regarding perceived personality factors and voice attractiveness. Additionally, a difference in the judgements of the listeners depending on gender was found: Female listeners rated the male voices significantly more positive than their male colleagues. That could be interpreted in terms of competition within gender and should be investigated more intensely in further research. These results, identifying the acoustic resources employed in the performance of dominance and possibly gender in general, are of great interest to sociolinguists and sociophoneticians.

## References

- Allport, G.W. and Cantril, H. 1934. Judging Personality from voice. *Journal of Social Psychology* 5, 37-55.
- Brown, B.L. 1982. Experimentelle Untersuchungen zur Personenwahrnehmung aufgrund vokaler Hinweisreize. In Scherer, K. (ed.), *Vokale Kommunikation*. Weinheim und Basel, Beltz Verlag, 211-227.
- Davis, S.B. 1979. Acoustic characteristics of normal and pathological voices. In Lass, N.J. (ed.), *Speech and language: Basic advances in research and practice*. New York, Academic Press, 271-335.
- de Saint-Exupéry, A. (2000). *Der kleine Prinz*. Karl Rauch-Verlag.
- Gélinas-Chebat, C., Chebat, J.-C. and Boivin, R. 2003. Voice and Information Processing. *Proc. of the 15th Int. Congress of Phonetic Sciences*, 667-670, Barcelona, Spain.
- Laver, J. 1994. *Principles of phonetics*. Cambridge, University Press.
- Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. 1957. *The measurement of meaning*. University of Illinois.
- Sendlmeier, W. and Siegmund, J. 2005. DeutschlandRadio Berlin vs. Radio NRJ Berlin - Ein Vergleich der Sprechstile. In Sendlmeier, W. (ed.) 2005, *Sprechwirkung. Sprechstile in Funk und Fernsehen*, 121-149. Berlin, Logos Verlag.
- Teshigawara, M. 2003. Voices in Japanese Animation: How People Perceive Voices of Good Guys and Bad Guys. *Proc. of the 15th Int. Congress of Phonetic Sciences*, 2413-2416, Barcelona, Spain.
- Walton, J.H. and Orlikoff, R.F. 1994. Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research* 37, 738-745.
- Yumoto, E., Gould, W.J. And Baer, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America* 71, 1544-1550.
- Zuckerman, M. and Driver, R.E. 1989. What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior* 13 (2), 67-82.
- Zuckerman, M. and Miyake, K. 1993. The attractive voice: What makes it so? *Journal of Nonverbal Behavior* 17, 119-135.

# Does uncertainty effect the case of exhaustive interpretation?

Charlotte Wollermann<sup>1</sup> and Bernhard Schröder<sup>2</sup>

<sup>1</sup>Institute of Communication Sciences, Department of Speech and Communication,  
University of Bonn, Germany

<sup>2</sup>German Linguistics, University of Essen/Duisburg, Germany

## Abstract

We present an experimental study investigating the role of intonation as prosodic indicator of uncertainty as well as question-answering congruity for the exhaustive interpretation of answers. We vary intonation for expressing intended (un)certainity and also the type of question. Interpretation is tested by using pictures illustrating (non-)exhaustivity in order to avoid that the subjects' linguistic awareness is focussed on the tested question.

Key words: exhaustive interpretation, focus detection, uncertainty, intonation, question-answering congruity

## Introduction

If the hearer concludes from (1b) that John and Mary are the only persons out of a number of people in question who passed the examination, the predicate in question in (1a) is interpreted exhaustively (cf. van Rooij, Schulz 2006: 205). In the case of non-exhaustive interpretation there are (or may be) also other persons who passed the examination.

- 1a      Who passed the examination?
- 1b      John and Mary.

According to semantic-pragmatic theories (e.g. Groenendijk, Stokhof 1984; Rooth 1992) in a context of a question, *accent* is highly correlated with focus. This should have a clear impact on exhaustive interpretation, especially in the context of a suitable question. An exhaustive interpretation is dependent on the knowledge about the situation in question, as ascribed to the speaker by the hearer. If the speaker is believed to have only uncertain knowledge of a situation, the hearer will less attract an exhaustive interpretation.

Several studies found that *rising intonation* is one of the acoustic cues which contribute to the perception of uncertainty (e.g. Smith, Clark 1993; Swerts et al. 2003). The analysis of Ward and Hirschberg (cf. 1985: 747) showed that *fall-rise* intonation contributes to a context-independent meaning of utterance interpretation conveying speaker's uncertainty<sup>1</sup>.

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.

However, there is barely empirical evidence of the role of uncertainty for *exhaustive interpretation*.

## **Experimental study**

### **Goal**

The goal of our experiment is to investigate whether intonation as prosodic indicator of uncertainty and the type of question effect exhaustive interpretation of answers.

### **Method**

Our audio stimuli consist of question-answer pairs, which are embedded into dialogs. Altogether there are six dialogs. The scenario is a fictional student dress-up party, where different student groups carry out different actions.

The focus of the answer is each time a noun phrase. It is either marked by *fall-rise* intonation (expressing uncertainty), or *falling* intonation (expressing certainty). The preceding question is either syntactically and lexically parallel to the answer or constitutes a more general question.

Our hypothesis is that falling intonation combined with parallel question favors exhaustive interpretation, whereas fall-rise intonation combined with general question advantages non-exhaustive interpretation. Independent variables are thus *intonation* and the *type of question*, the dependent variable is *exhaustivity*.

For testing the interpretation, subjects have to choose between different pictures<sup>2</sup>. One picture illustrating exhaustive interpretation, another showing non-exhaustive reading and a third picture showing an irrelevant scene functioning as distractor. Each group is characterized by different accessories, which is illustrated by the pictures. Also as distractor, we ask questions about the subjects' personal opinion of an aspect of the dialog and use two filler-dialogs<sup>3</sup>.

Subjects are 71 students (24 males, 47 females) of the Institute of Communication Sciences, all native speakers of German. They are tested in four group experiments: Subsets of the dialogs are presented each time with a different kind of random order of the stimuli and also of the pictures. We use Fisher's exact test to examine the significance of the association between two categorical variables.

### **Results**

Results are presented in Table 1. As the table shows most interpretations are exclusively exhaustive, but there are also some non-exhaustive interpretations. The number varies according to the combinations and the dialogs. Significant differences are as follows: For Dialog 2 the comparison



of general question combined with certainty (C) and general question plus uncertainty (D) yields a significant difference with  $p=0.01$ . Here, intonation seems to clearly influence the judgments.

For Dialog 4 general question plus certainty (C) is judged as more non-exhaustive than certainty combined with congruent question (A) ( $p=0.02$ ). Thus, question-answering congruity favors exhaustivity in this case. In accordance, general question joined with uncertainty (D) obtains more non-exhaustive readings than congruent question plus certainty (A) with  $p=0.049$ .

Table 1. Amount of answers in both categories of exhaustivity. Abbreviations: *Exh+*: Exclusively exhaustive interpretations, *Exh-*: Readings that comprehend non-exhaustive interpretation, *CQ*: Congruent question, *GQ*: General question, *C*: Certainty, *U*: Uncertainty. Significant different judgments are marked in bold face.

Stimulus	Description	D1		D2		D3		D4	
	Exh.	+	-	+	-	+	-	+	-
A	CQ+C	14	4	17	2	9	6	15	0
B	CQ+U	12	3	16	3	17	2	17	1
C	GQ+C	12	6	18	0	12	7	9	5
D	GQ+U	14	9	10	5	16	2	13	5

Stimulus	Description	D5		D6	
	Exh.	+	-	+	-
A	CQ+C	16	3	15	4
B	CQ+U	8	7	16	2
C	GQ+C	13	6	16	2
D	GQ+U	16	2	10	5

For Dialog 5 we have more non-exhaustive interpretations for congruent question joined with uncertainty (B) than with certainty (A). Differences are marginally significant ( $p=0.068$ ). We assume that intonation influences exhaustivity here. But surprisingly, general question with uncertainty (D) is ranked as more exhaustive than (B) with  $p=0.047$ . Here, the tendency does not mirror what was expected theoretically. This could possibly be explained by the fact that in this single case the fall-rise intonation is realised more conspicuous than the falling intonation and thus favors exhaustivity.

We assume that for Dialog 2, 4 and 5 as our three “successful” dialogs, intonation and linguistic context (i.e. choice of the verbs) is realised more adequate than for the remaining three “unsuccessful” dialogs.



## Discussion

The experiment brought to light that exhaustive reading is considered as standard interpretation in our scenario. However, in three of six dialogs, results suggest that exhaustivity can be influenced by question-answering congruity and also by intonation. But our data reveal that intonation exclusively does not have such a strong effect on exhaustivity as semantic-pragmatic theories would suggest.

Future work will concentrate on the “successful” dialogs. We will investigate if additional prosodic cues indicating uncertainty like pauses and a hesitant way of speaking combined with fall-rise intonation and/or contextual factors facilitate non-exhaustive interpretation.

## Notes

- <sup>1</sup> *Fall-rise* intonation is defined as follows: Firstly, the pitch peak is reached late in the accented syllable and a relatively abrupt drop in pitch must appear in the two following syllables. Secondly, a sentence-final rise in pitch is at hand (cf. Ward, Hirschberg 1985: 748).
- <sup>2</sup> For each dialog, the pictures and questions are presented on slides and presented via a beamer, subjects only have to mark their choice on a handed-out questionnaire. In this way, we want to assure that subjects do not page back and change their judgments.
- <sup>3</sup> Subjects have the possibility to choose multiple answers for both tasks.

## Acknowledgements

We would like to thank Petra Wagner, Ulrich Schade, Eva Lasarczyk, and Bernhard Fisseni for helpful discussions and comments. Many thanks go to our speakers Julia Abresch and Stefan Keller.

## References

- Groenendijk, J. and Stokhof, M. 1984. Studies on the Semantics of Questions and the Pragmatics of Answers. Dissertation, University of Amsterdam.
- Rooth, M. 1992. A theory of focus interpretation. In *Natural Language Semantics*, 75-116.
- Smith, V. and Clark, H. 1993. On the course of answering questions. In *Journal of Memory and Language*, 32, 25-38.
- Swerts, M., Krahmer, E., Barkhuysen, P. and van de Laar, L. 2003. Audiovisual cues to uncertainty. In *Proceedings of ISCA Workshop on Error Handling in Spoken Dialog Systems*, Chateau-d'Oex, Switzerland.
- van Rooij, R. and Schulz, K. 2006. Pragmatic Meaning and Non-monotonic Reasoning: The Case of Exhaustive Interpretation. In *Linguistics and Philosophy*, 29, 205-250.
- Ward, G. and Hirschberg, J. 1985. Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation. In *Language*, 61(4), 747-776.

## **“Sounds like a rainbow” - sound-colour mappings in vowel perception**

Magdalena Wrembel and Karolina Rataj

School of English, Adam Mickiewicz University, Poland

### **Abstract**

The paper reports on an experiment conducted to investigate the nature of speech sound perception in terms of cross-modal mappings between vowel sound stimuli and colour spectrum associations. The study is based on the assumptions stemming from research on synaesthesia, sound symbolism and non-modularity of human perception. The findings indicate that vowel-sound mappings appear non-arbitrary in non-synaesthetic perception and follow the general tendencies in which bright colours are associated with prominent high-pitched sounds, whereas dark colours are attributed to lower-pitched tones. The results may have implications for L2 pronunciation pedagogy in that they may enhance the effectiveness of L2 phonological acquisition.

Key words: sound-colour mappings, vowel colour, synaesthesia

### **Theoretical assumptions**

The application of colour terminology with relation to vowels can be traced back to Jakobson (1962) who pointed to the regularity of colour associations in coloured hearing synaesthesia identifying close connections of vowels /o/ /u/ with darker colours, /e/ /i/ with brighter colours, and /a/ with red. However, very few studies to date have explored the phenomenon of making associations between colour spectrum and speech sounds in normal perception. Flagg and Stewart (1985) claimed that 6 primary colours can be used to facilitate consonantal perception. Dailey et al. (1997) investigated the relation between creativity, synaesthetic tendencies and physiognomic perception. Their findings demonstrated that creative individuals have access to primary thinking processes that assume a unity of different sensory modalities since they exhibited stronger associations between colours and vowels.

Donegan (1985) used the term *vowel colour* to refer to such features as palatality and labiality. She stated that palatal vowels, traditionally referred to as front vowels, which have high F2 and a considerable distance between F1 and F2, tend to be perceived as 'bright' as opposed to 'dark' labial vowels characterised by low F2 and a small distance between F1 and F2. Vowels which are neither palatal or labial are called plain or achromatic.

Finally, researchers report a strong correlation between auditory pitch and visual luminance; there is a general tendency for most people to associate

high pitch sounds with light colours and low tones with dark colours (e.g. Hubbard 1996, Ward et al. 2006). As suggested by Ward et al. (2006), research on synaesthesia can be used “to inform theories of normal cognition”, and further verification is needed of the hypothesis he proposes, which states that mechanisms analogous to the ones used in synaesthetic perception may be recruited in non-synaesthetic cross-modal perception.

## **Experiment**

The aim of the present study was to investigate whether L2 speech sounds evoke associations with specific colours in non-synaesthetic speakers highly proficient in English.

## **Participants and materials**

48 first year students of English at Adam Mickiewicz University in Poznań, Poland took part in the experiment. The participants consisted of 34 females and 14 males and their mean age was 20 years. At the time of the experiment they were ignorant about the phenomenon of synaesthesia and an interview conducted prior to the experiment served to rule out this ability in the tested population.

Sound stimuli used in the experiment included 12 English pure vowel sounds recorded in isolation. The stimuli were recorded by a male native speaker of English (in a professional recording studio) as 16-bit mono files at a sampling frequency of 16000 Hz using Audacity software.

## **Procedure**

The experiment was run on a specially designed computer program, implemented in Visual Basic, that offered the following functionality: (1) playing randomised sounds; (2) displaying the palette of 11 basic colours in a randomised order; (3) registering participants' data and responses. The experiment was carried as a series of individual sessions. The participants were seated in front of a computer screen in a dimmed room. They were instructed to listen to individual sounds and choose one colour from a palette of 11 basic colours, as specified by Berlin and Kay (1969). The test was self-paced and the participants could play a given sound as many times as they wished but once a colour was selected they were not able to modify it. The choice was made by clicking on one of the 11 coloured rectangles presented in 3 rows against a light grey background. The colour palette appeared automatically on the screen with a 2s delay after a 'Play' button was clicked in order to play a particular vowel sound. The whole procedure was repeated after a 7-week interval and yielded a total of 75 sets of responses.

## Results and discussion

The SPSS analysis revealed the statistical significance of sound-colour associations for 8 out of 12 English vowels under investigation ( $p < .01$ ). See Table 1 for the results of a nonparametric chi-square test.

Table 2 presents the data generated by the two parts of the experiment ( $N=75$ ), highlighting in yellow these particular colour associations which were above chance level as demonstrated by the chi-square test results.

Table 1. Chi-square analysis for vowel-colour associations in English.

	æ	ɔ:	ʌ	ɑ:	e	ɜ:	ɪ	i:	ɒ	ə	ʊ	u:
chi2	26.6	41.9	14.3	15.5	26.9	40.5	17.5	58	10.2	30.4	50.4	23
df	10	9	10	10	10	10	10	10	10	10	10	9
p	.003	.000	.16	.115	.003	.000	.063	.000	.422	.001	.000	.006

Table 2. Vowel-colour associations for English (highlighted values statistically significant at  $p < .01$ ).

N=75	æ	ɔ:	ʌ	ɑ:	e	ɜ:	ɪ	i:	ɒ	ə	ʊ	u:
grey	7%	11%	15%	3%	4%	20%	17%	3%	11%	23%	28%	7%
brown	5%	29%	4%	9%	4%	19%	7%	1%	4%	8%	17%	24%
yellow	17%	7%	14%	8%	7%	3%	11%	28%	11%	8%	3%	5%
green	5%	8%	14%	7%	16%	7%	11%	20%	4%	13%	8%	8%
orange	11%	4%	11%	7%	15%	1%	16%	9%	12%	13%	9%	11%
black	11%	16%	1%	16%	5%	16%	4%	1%	11%	3%	13%	13%
blue	4%	12%	9%	11%	12%	9%	7%	8%	13%	7%	4%	13%
red	21%	5%	8%	15%	19%	4%	4%	5%	11%	3%	3%	7%
purple	9%	0%	7%	11%	7%	14%	12%	7%	8%	4%	7%	8%
white	4%	4%	11%	12%	1%	1%	5%	13%	12%	13%	4%	0%
pink	5%	4%	7%	3%	8%	5%	7%	4%	4%	5%	4%	4%

Table 3. Statistically significant colour associations for vowel categories.

	Front	Central	Back
yellow	16%		
green	13%		
orange	13%		
grey		19%	14%
brown			19%

	High	Mid	Low
red			15%
grey	14%	13% *	

$p < .001$

Strong correlations were also found when the analysis accounted not for individual sounds but rather vowel categories (front/central/back and

high/mid/low). The assignment of colours was statistically significant in 5 out of 6 categories and in one it approached significance (marked with \*, see Table 3).

The present analysis relied mostly on local characteristics of formant frequencies as perceptual parameters used to categorise vowel colour. The findings indicate that vowel-sound mappings in non-synaesthetic perception appear non-arbitrary and follow the general tendencies in which bright colours (yellow, green) are associated with prominent high front vowel sounds, whereas dark colours (brown, blue, black) are attributed to back vowels, open sounds tend to be perceived as red and central vowels are mapped onto achromatic grey. Thus the results seem to corroborate the hypothesis that similar mechanisms may be used in non-synaesthetic and synaesthetic perception of colours and sounds as previously indicated by Ward et al. 2006, however, these mappings cannot be expected to equal the consistency and specificity of colours demonstrated by synaesthetes.

Further research into this phenomenon is definitely needed to verify the hypothesis more thoroughly. Moreover, it may provide specific implications for L2 pronunciation pedagogy and thus may enhance the effectiveness of second language phonological acquisition.

## References

- Berlin, B., Kay, P. 1969. Basic colour terms: Their universality and evolution. Berkeley, CA: University of California Press.
- Dailey, A., Martindale, C., Borkum, J. 1997. Creativity, Synaesthesia, and Physiognomic Perception. *Creativity Research Journal*, 10: 1-8.
- Donegan, P. 1985. *On the Natural Phonology of Vowels*. New York: Garland Publishing.
- Flagg, L., Stewart, J. 1985. Studying speech perception in adolescent school-age children by primary color perception. *J. of Psychol. Research*, 14: 67-80.
- Hubbard, T.L. 1996. Synesthesia-like mappings of lightness, pitch and melodic interval. *American Journal of Psychology*, 109: 219-238.
- Jakobson, R. 1962. *Selected Writings: I Phonological Studies*. The Hague: Mouton.
- Ward, J., Huckstep, B., Tsakanikos, E. 2006. Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all?, *Cortex*, 42, 264-280.

# Focus effects on syllable duration in Cypriot Greek

Charalabos Themistocleous

Department of Linguistics, University of Athens, Greece

## Abstract

The present experimental study examined the effects of focus on the duration of stressed syllable onset and rhyme in words found in prefocal, focal, post focal and neutral position. The main results generated indicate a significant effect of focus position on the segmental duration. Additionally, word final lengthening was demonstrated, whereas no word initial lengthening effects were observed. Furthermore, the results showed rightward lengthening effects and leftward shortening effects due to focus position on stressed syllables.

Keywords: duration, stress, focus, prosody.

## Introduction

This experimental study examines the effects of focus position on the duration of stressed syllable onset and rhyme in Cypriot Greek (henceforth CG) regarding the following questions: (a) What are the effects of focus position on stress segmental duration? (b) Does focus affects the duration of all syllables within a word in the same way? These questions address critical phonological considerations concerning duration as a means of marking focal domains (c.f. Beckman and Pierrehumbert 1986), as well as considerations about word initial and final lengthening (c.f. Botinis et al. 2001; Turk and Shattuck-Hufnagel 2000).

## Methodology

The target syllable was the stressed [ˈla] in antepenultimate, penultimate and ultimate position. Three keywords were chosen, with three CV syllables each, with one of them being the target syllable [ˈla]: [ˈla.pi.θɔs] ‘Lapithos’, [pi.ˈla.ðis] ‘Pyladis’ and [ma.ju.ˈla] ‘Majula’. These were uttered in the carrier-sentence: [i ˈelli ˈlei\_ˈlaθɔs] ‘Elli says \_ erroneously’, in four different focus positions: preceding focus, focus on keyword, following focus and neutral focus. The materials were recorded by six native speakers of CG, in their early twenties. Each speaker had to utter 4 sentences x 3 keywords x 10 repetitions. The total corpus consisted of 720 utterances. The sentences were typed in Greek orthography and each prompt was presented in random order before a subject. The test words were manually segmented and labelled, by using simultaneous inspections of waveforms and wide-band spectrograms following standard criteria (e.g. Peterson and Lehiste 1960). The durations of the syllable [la], the syllable onset [l] and the rhyme [a] were measured.

---

Proceedings of the 2<sup>nd</sup> ISCA Workshop on Experimental Linguistics, ExLing 2008, 25-27 August 2008, Athens, Greece.

## Results

Statistical analysis was carried out and the mean durations of syllables in phrases uttered with four distinct foci are shown in Figure 1:

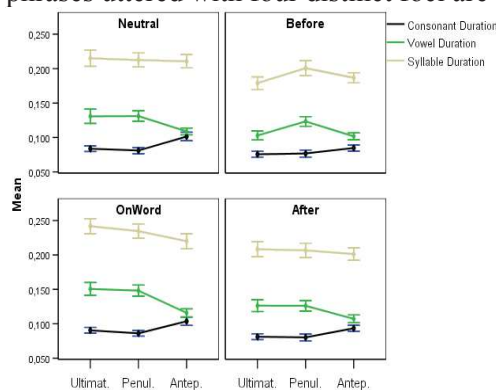


Figure 1. Mean consonant, vowel & syllable duration in seconds in ultimate, penultimate and antepenultimate position, uttered in words with four distinct focus positions i.e. with neutral focus (Neutral), with focus accent preceding word (Before), with focus on the word (OnWord) and with focus accent following word (After); error bars show Std Dev.

The correlations of FOCUS X SYLLABLE POSITION are shown for consonant, vowel and syllable duration. A two way ANOVA was performed; all effects are reported at a .05 level of significance.

(a) *Consonant Duration*: Consonants in syllables under focus were longer than in phrases with other focus patterns. Specifically, consonants in antepenultimates under focus ( $M=.104$ ) were longer than consonants in penultimates ( $M=.087$ ) and ultimates ( $M=.090$ ); consonants in syllables with neutral focus followed ( $M=.089$ ). Interestingly, the shortest consonants were in penultimates with preceding focus ( $M=.077$ ). Levene's test indicated that the assumption of homogeneity of variance had been violated,  $F(11, 708)=1.993, p < .05$ , however  $F$ -tests are reported. The results showed that *syllable position* significantly affected the produced consonant duration,  $F(2, 708)=46.382, p < .05, r = 1$ . *Focus position* affected consonant duration,  $F(3, 708)=19.478, p < .05, r = 1$ . *Focus position x syllable position* interaction was not significant,  $F(6, 708)=1.383, ns$ . Games-Howell post hoc test for syllable position showed significant differences between all groups ( $p < .05$ , in all cases), except between ultimates and penultimates,  $ns$  and for focus position between all groups ( $p < .05$ , in all cases), except between neutral focus and focus following word, and between neutral and focus on word,  $ns$ .

(b) *Vowel Duration*: Vowels in syllables under focus were longer than in phrases with other focus patterns (see Figure 1). Specifically, vowels in *ultimates* under focus ( $M=.151$ ) were longer than vowels in *penultimates* ( $M=.148$ ) and *antepenultimates* ( $M=.116$ ). Vowels in syllables with neutral focus follow in mean duration ( $M=.124$ ). Interestingly, the shortest vowels were in syllables with preceding focus, especially the vowels found in antepenultimates ( $M=.102$ ) and ultimates ( $M=.103$ ). Levene's test



indicated that the assumption of homogeneity of variance had been violated,  $F(11, 708) = 4.204, p < .05$ , however  $F$ -tests are reported. Results showed that *syllable position* significantly affected vowel duration,  $F(2, 708) = 46.889, p < .05, r = .37$ . *Focus position* significantly affected vowel duration,  $F(3, 708) = 31.392, p < .05, r = .35$ . The *focus position x syllable position* interaction was significant,  $F(6, 708) = 4.159, p < .05, r = .20$ . Games-Howell post hoc tests for syllable position revealed significant differences between all groups ( $p < .05$ ), except between ultimates and penultimates, *ns*. For focus position significant differences were revealed between all groups ( $p < .05$ ) except between neutral focus and following focus, and between neutral and focus on word *ns*.

(c) *Syllable Duration*: The duration of syllables in words under focus was longer than in phrases with other focus patterns, especially the duration of ultimates under focus ( $M = .242$ ) was longer than of penultimates ( $M = .235$ ) and antepenultimates and ( $M = .220$ ). Syllables with neutral focus follow in mean duration ( $M = .213$ ). Interestingly, the shortest syllables were in words with preceding focus, especially antepenultimates ( $M = .187$ ) and ultimates ( $M = .179$ ) following the vowels' pattern. The results showed that the main effect of the type of *syllable position* in the produced duration of syllables significantly affected syllable duration,  $F(2, 708) = 3.338, p < .05, r = .05$ . *Focus Position* significantly affected syllable duration,  $F(3, 708) = 37.542, p < .05, r = .36$ . *Focus position x syllable position* interaction was significant,  $F(6, 708) = 2.314, p < .05, r = .12$ . Games-Howell post hoc tests for syllable position within the word revealed significant differences between penultimate and antepenultimate ( $p < .05$ ); all other cases were non-significant *ns*. There were significant effects of focus position in all groups, ( $p < .05$ ), except between neutral focus and focus following word *ns*.

## Discussion

Results generated support that focus position significantly affects the duration of stressed syllables. Observations on these data, without any accounts of focus position revealed no *word final lengthening* (see also Turk and Shattuck-Hufnagel 2000; Katsika 2007), attributed to shortening effects of preceding focus accent, most evident on ultimate syllables. As a result penultimates had increased duration comparing to ultimates and antepenultimates. A striking result was the effect of focus position with regards to *word final lengthening*: When focus was accounted for in the model, stressed ultimate syllables aligning with focus accent showed increased duration comparing to ultimates with preceding focus, which were even shorter than neutral accented syllables. Additionally, stressed antepenultimates and stressed penultimates had a constant relation: *syllables with focus > syllables with neutral focus > syllables with following focus >*



*syllables with preceding focus accent*. Furthermore, stressed syllables had various lengthening due to focus: ultimates were more prone to lengthening or shortening than antepenultimates and penultimates. Although these observations are in accordance with previous studies (c.f. Klatt 1976, Cooper and Paccia-Cooper 1980, Botinis 1989), the small effect size of the tests, along with counter evidence from studies on Athenian Greek (AG) (c.f. Botinis et al. 2001) suggest that these results should be considered with caution. Critical observations were reported for the sub-syllable level (see Turk and Shattuck-Hufnagel 2000): (a) word final lengthening is localized on the ‘*rime*’ of the final syllable, *the vowel* and (b) word initial lengthening is localized on the onset of the initial syllable, *the consonant*. Even though, the first prediction was confirmed by the data, the second was not; *initial word lengthening* was not supported by the results of the present study as word initial stressed syllables were found to be shorter in most cases than ultimate and penultimate syllables, despite the fact that the consonants of these syllables were evidently longer. This observation, however, contrasts previous studies on AG (see Botinis et al. 2001, Katsika 2007) that support initial lengthening. However these studies examine AG, while the present one is based on CG data, therefore additional comparative evidence is needed. Furthermore, this study did not take into account the effects of speech tempo. Evidence for the effects of tempo in Greek has been observed by Botinis et al. (2001). Also, comparative measurements from unstressed syllables are essential in assisting the interpretation of these results.

## References

- Beckman, M. and Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255-309.
- Botinis, A. 1989. *Stress and Prosodic Structure in Greek*. Lund: Lund University Press.
- Botinis, A., Fourakis, M. and Bannert, R. 2001. Prosodic interactions on segmental durations in Greek. *Lund University, Working Papers* 49, 10–13.
- Cooper, W.E. and Paccia-Cooper, J. 1980. *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Katsika, A. 2007. Duration and pitch anchoring as cues to word boundaries in Greek. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 929-932. Saarbrücken: Universität des Saarlandes.
- Klatt, D. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Amer.* 59(3), 1208-1221.
- Peterson, G.E. and Lehiste, I. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Amer.* 32, 693–703.
- Turk, A.E. and Shattuck-Hufnagel, S. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, 397-440.
- Turk, A.E. and White, 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27, 171-206.

## Index of authors

---

- Alagözlü, N., 1  
Alenazi, A., 9  
Alexandris, Ch., 5  
Alghamdi, M., 9  
Alhargan, F., 9  
Alkanhal, M., 9  
Alkhairy, A., 9  
AlShammari, E.T., 13  
Andreou<sup>a</sup> G., 17  
Anufryk, V., 21  
Arai, T., 161  
Arnold, D., 25  
Arts, A., 29  
Augurzky, P., 33  
Bakamidis, S., 117  
Baumotte, H., 37  
Bijankhan, M., 153  
Botinis, A., 41  
Boucher, V.J., 46  
Campos-Astorkiza, R., 50  
Cantoni, M., 54  
Castro<sup>a</sup> L., 58, 198  
Chaida, A., 62  
Chentir, A., 66  
Chu, M.-N., 70  
Correia, D., 74  
de Moraes, J.A., 58  
den Ouden, H., 78  
Dogil, G., 21, 37, 138  
Eldesouki, M., 9  
Falé, I., 82, 86  
Faria, I.H., 74, 86, 90  
Fernández-Parra, M., 94  
Fourakis, M., 41  
Galantomos, I., 17  
Gawronska, B., 106  
Gennari, S.P., 146  
Gilbert, A.C., 46  
Giordano, R., 98  
Gomes, Ch.A., 206  
Guerti, M., 66, 178  
Gussenhoven, C., 70  
Hansakunbuntheung, Ch., 102  
Hemerén, P.E., 106  
Herrero, B.P., 110  
Hirst, D., 66  
Jannedy, S., 114  
Jansen, C., 29  
Jilka, M., 21  
Kalimeris, C., 17  
Rataj, K., 238  
Kasviki, S., 106  
Kato, H., 102  
Kinoshita, N., 122  
Kocjancic, T., 126  
Lehtinen, M., 130  
Lengeris, A., 134  
Lewandowski, N., 138  
Lin, J., 13  
Liu, Ch.-J., 141  
Liu, H.-T., 141  
Loui, S., 145  
Loukina, A., 149  
Luegi, P., 73, 89  
Maes, A., 29  
Mahmoodzade, Z., 153  
Martin, Ph., 157  
Masuda, H., 161  
Melinger, A., 225  
Mitterer, H., 165  
Mompean, J.A., 169  
Moosmüller, S., 173  
Nikolaenkova, O., 41  
Noordman, L., 29  
Ouamour, S., 177  
Paraskevas, I., 181  
Pastuszek-Lipińska, B., 185  
Rangoussi, M., 181  
Russo, M., 189  
Sadeghi, V., 193  
Sagisaka, Y., 101, 217  
Sayoud, H., 177  
Schröder, B., 233  
Serridge, B., 197  
Sheppard, Ch., 201  
Silva, Th.C., 205  
Sonu, M., 209  
Themistocleous, Ch., 241  
Todaka, Y., 213  
Tseng, Ch.-H., 141  
Tsubaki, H., 217  
Wade, T., 137  
Wagner, A., 221  
Wagner, P., 25  
van Hout, R., 69  
van Wijk, C., 77  
Weber, A., 225  
Weirich, M., 229  
Wollermann, Ch., 233  
Wrembel, M., 237  
Yamamoto, H., 217  
Yasuda, K., 217

ΤΥΠΩΘΗΚΕ ΣΤΟ Ε.Κ.Π.Α.  
Σταδίου 5, Τ.Κ. 105 62 - Αθήνα,  
Τηλ. 210 36.89.374-210 36.89.375-210 36.89.388,  
Fax: 210 36.89.433  
e-mail: [publish@elke.voa.gr](mailto:publish@elke.voa.gr)



**ISBN: 978-960-466-020-9**