the Phonetician

A Publication of ISPhS/International Society of Phonetic Sciences

NUMBER 104-105 2012/I-II

ISPhS
International Society of Phonetic Sciences

President: Ruth Huntley Bahr

Secretary General:
Mária Gósy

Honorary President:
Harry Hollien

Vice Presidents:
Angelika Braun
Marie Dohalská-Zichová
Mária Gósy
Damir Horga
Heinrich Kelz
Stephen Lambacher
Asher Laufer
Judith Rosenhouse

Past Presidents:
Jens-Peter Köster
Harry Hollien
William A. Sakow †
Martin Kloster-Jensen†
Milan Romportl †
Bertil Malmberg †
Eberhard Zwirner †
Daniel Jones †

Honorary Vice Presidents:

A. Abramson	P. Janota	H. Morioka	M. Shirt
S. Agrawal	W. Jassem	R. Nasr	E. Stock
L. Bondarko	M. Kohno	T. Nikolayeva	M. Tatham
E. Emerit	E.-M. Krech	R. K. Potapova	F. Weingartner
G. Fant	A. Marchal	M. Rossi	R. Weiss

Auditor:
Angelika Braun

Treasurer:
Ruth Huntley Bahr

Affiliated Members (Associations):

American Association of Phonetic Sciences	J. Hoit & W.S. Brown
Dutch Society of Phonetics	B. Schouten
International Association for Forensic Phonetics and Acoustics	A. Braun
Phonetic Society of Japan	I. Oshima & K. Maekawa
Polish Phonetics Association	G. Demenko

Affiliated Members (Institutes and Companies):

KayPENTAX, Lincoln Park, NJ, USA	J. Crump
Inst. for Advanced Study of the Communication Processes, University of Florida, USA	H. Hollien
Dept. of Phonetics, University of Trier, Germany	J.-P. Köster
Dept. of Phonetics, University of Helsinki, Finland	A. Iivonen
Dept. of Phonetics, University of Zürich, Switzerland	S. Schmid
Centre of Poetics and Phonetics, University of Geneva, Switzerland	S. Vater

International Society of Phonetic Sciences (ISPhS) Addresses

www.isphs.org

President:

Professor Ruth Huntley Bahr, Ph.D.
President's Office:
Dept. of Communication Sciences
and Disorders
University of South Florida

4202 E. Fowler Ave., PCD 1017
Tampa, FL 33620-8200
USA
Tel.: ++1-813-974-3182
Fax: ++1-813-974-0822
e-mail: rbahr@usf.edu

Secretary General:

Prof. Dr. Mária Gósy
Secretary General's Office:
Kempelen Farkas Speech
Research Laboratory,
Research Institute for Linguistics,
Hungarian Academy of Sciences
Benczúr u. 33
H-1068 Budapest
Hungary
++36 (1) 321-4830 ext. 172
++36 (1) 322-9297
e-mail: gosy.maria@nytud.mta.hu

Guest Editors:
Steven M. Lulich, Ph.D.
Guest Editor's Office:
Department of Speech and Hearing Sciences
Indiana University
200 S. Jordan Avenue
Bloomington, Indiana 47405
USA
Tel.: +1 (812) 856 2423
Email: slulich@indiana.edu

Review Editor:
Prof. Judith Rosenhouse, Ph.D.
Review Editor's Office:
Swantech
89 Hagalil St
Haifa 32684
Israel
Tel.: ++972-4-8235546
Fax: ++972-4-8327399
e-mail: swantech@013.net.il

Tekla Etelka Gráczi
Guest Editor's Office:
Kempelen Farkas Speech Research
Laboratory,
Research Institute for Linguistics,
Hungarian Academy of Sciences
Benczúr u. 33
H-1068 Budapest
Hungary
Tel.: ++36 (1) 321-4830 ext. 172
Fax: ++36 (1) 322-9297
e-mail: graczi.tekla.etelka@nytud.mta.hu

Cover: *A concert* by Lorenzo Costa (1459/60-1535)
1485-95, oil on wood, 95.3 x 75.6 cm
National Gallery, London

INTRODUCING THE GUEST EDITORS

Dr. Steven M. Lulich has been a research scientist and occasional lecturer in the departments of Speech & Hearing Science and Linguistics at Indiana University, Bloomington, and will be starting a new lab as an assistant professor in the Speech & Hearing Science department in the Fall of 2013. Prior to coming to Bloomington, he was a research scientist in Professor Mitchell S. Sommers' lab in the Psychology Department at Washington University in St. Louis, following a post-doc in Molecular and Integrative Physiology in Professor Jeffrey J. Fredberg's lab at the Harvard School of Public Health. As an undergraduate he studied Linguistics and Classical Greek with Professsors Ioana Chitoran and Lindsay Whaley at Dartmouth College (BA cum laude, with honors), and was a foreign exchange student for one year with Professors Grzegorz Dogil and Bernd Möbius at the Institute for Natural Language Processing at the University of Stuttgart, Germany. He earned his PhD in Speech & Hearing Bioscience & Technology from MIT in 2006 under the direction of Professor Kenneth N. Stevens, and has taught at MIT, the Budapest University of Technology and Economics, and Indiana University. His research interests include lung, speech and voice acoustics, speech anatomy and physiology, articulatory-acoustic relations and inversion, speech signal processing, automatic speech recognition and text-to-speech synthesis, and talker-specific models of speech production.

Tekla Etelka Gráczi is a junior research fellow at the Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary. She studied Hungarian Linguistics and Literature at Eötvös Loránd University (ELTE), Budapest Hungary. Her undergraduate thesis was written on the perception of obstruents and vowels in whispered speech under supervision by Prof. Dr. Mária Gósy. She did her PhD-studies in Hungarian Linguistics also at ELTE, and specialized on Phonetics. She defended her PhD-thesis on the acoustic properties of the voicing opposition of obstruents in Hungarian in January 2013, also under Prof Dr. Mária Gósy's supervision. She gave lectures at ELTE between 2007 and 2011. Her main research intrests are the acoustic and perceptual characteristics of voicing opposition, voice disorders, segmental variability amoung speech styles.

the Phonetician

A Publication of ISPhS/**International Society of Phonetic Sciences**

ISSN 0741-6164

Number 104-105 / 2012-I-II

CONTENTS

BOUNDARY MARKING IN HUNGARIAN V(#)V CLUSTERS WITH SPECIAL REGARD TO THE ROLE OF IRREGULAR PHONATION

Alexandra Markó
Department of Phonetics, Eötvös Loránd University, Budapest
e-mail: marko.alexandra@btk.elte.hu

Abstract

In the present paper, the realization of vowel clusters in Hungarian speech is analyzed. We focus our attention on cases in which the speaker wishes to highlight, rather than resolve, a hiatus – by employing irregular phonation. Glottalization occurred the most frequently (31.1%) across word boundaries; sometimes (with a frequency below 10%). It also happened morpheme internally or across compound boundaries. Glottalized word transitions were realized mostly at phrase boundaries (stress also influenced the occurrence of glottalization). Another major motivation for a glottalized realization of V(#)V clusters was to avoid the use of some phonological/articulatory mechanism (hiatus resolution or deletion). A large amount of inter-speaker variance was shown by the frequency of occurrence of glottalization.

1 Introduction

Hiatus is a heterosyllabic sequence of adjacent vowels, which is a dispreferred configuration in many languages (Siptár, 2002: 85). "Some languages disallow the occurrence of hiatus altogether; others prevent some instances from arising by various means but let others surface or resolve them in some surface-phonological manner" (Siptár, 2008: 189). The diverse means of avoiding hiatus include (1) the elision of one or the other vowel (e.g. *barn**a*** 'brown' + *ít* > *barnít* 'embrown', *kocsi* 'car' + ***o**n* > *kocsin* 'by car'), (2) glide formation: the change of one or the other vowel into a glide (e.g. *kale**i**doszkóp* > *kale*[j]*doszkóp* 'kaleidoscope'), and (3) "hiatus resolution", i.e., the spread of some segmental content from one of the flanking vowel positions into the empty onset position (e.g. *dió* > *di*[j]*ó* 'walnut') (Siptár, 2012).

In present-day Hungarian, regular hiatus resolution involves the consonant *j*, whose occurrence depends on the quality of the vowels constituting the cluster. Hiatus is invariably resolved if one of the two vowels concerned is *i* [i] or *í* [iː] (*ki*[j]*áltás* 'a cry', *si*[j]*et* 'hurry', *nő*[j]*i* 'woman-adj.'), and it is also often resolved if one of the vowels is *é* [eː] (if the *é* comes first, *j*-insertion is optional: *po*[j]*én* 'punch line' vs. *melléáll* [mɛlːeːaːlː] 'stand by' – for the exact criteria of this, e.g. Siptár and Törkenczy, 2000: 283–284). In other cases, hiatus resolution by [j] is not likely,

although individual speakers' habits might vary in terms of surface realization (e.g. *tea* 'tea': [tɛɔ] or [tɛjɔ]). When both vowels are low and/or rounded, hiatus resolution does not normally take place (e.g. *fáraó* [fa:rɑo:] 'pharaoh', *neon* [nɛon] 'neon', *műút* [my:u:t] 'highway'). Whether or not the two vowels are separated by a morpheme boundary has no bearing on the probability of hiatus resolution (Siptár and Törkenczy, 2000: 283).

In the present paper, we analyze the realization of vowel clusters in Hungarian speech in a perspective that has seldom been applied to this issue so far. In what follows, in addition to a general overview of the realizations of VV sequences, we will focus our attention on cases in which the speaker wishes to highlight, rather than resolve, a hiatus. We assume that one of the ways to do this may be to employ irregular phonation.

Modal voice is defined in the literature as quasi-periodic vibration (e.g. Gósy, 2004). However, in some cases, voice production may depart from this, and phonation may become irregular (glottalization, creaky voice).

It is not always easy to tell whether the quality of voicing is regular or otherwise: there are no agreed-on threshold values in the literature (e.g. in terms of jitter and/or shimmer) above which voice production can be said to be irregular. Furthermore, irregularity can show up in a number of forms. For instance, Batliner et al. (1993) distinguished six types of irregular phonation (laryngealization) in approx. 30 minutes of spontaneous and read speech by four speakers. Dilley et al. (1996) studied texts read aloud by five speakers in radio news programs with respect to irregular voice quality occurring in word initial vowels. They defined four types of realizations.

Glottalization is a multifunctional phenomenon. In some languages, it expresses a phonological contrast – mostly it distinguishes pairs of sonorants from one another (for instance, in Mazateco, spoken in Mexico, it distinguishes vowels, and in some North-American Indian languages it distinguishes nasals); less frequently (e.g. in Hausa) distinguishing obstruents (Gordon and Ladefoged, 2001). In several dialects of English, glottalization distinguishes allophones of syllable-final /t/ and /p/ (Pierrehumbert and Talkin, 1992).

Several researchers have investigated the role of glottalization in expressing emotions and/or tried to use it in the automatic recognition of emotions (e.g. Batliner, et al. 2007; Gobl and Ní Chasaide, 2003). The socio-cultural role of glottalization has also been demonstrated in an experiment involving young American women (Yuasa 2010), and its conversational function has also been supported for English, where the realization of *yeah* with modal vs. irregular phonation is associated with distinct functions by the speaker (and the listener) (Grivičić and Nilep, 2004).

The frequency of occurrence of glottalization is speaker dependent to a large extent: some speakers produce irregular voicing hardly at all, while some produce it fairly frequently (Henton and Bladon, 1988; Dilley et al., 1996; Redi and Shattuck-Hufnagel, 2001; Slifka, 2006; for Hungarian: Markó, 2005; Bőhm and Ujváry,

2008). Therefore, voice quality has an eminent role in human speaker recognition (Bőhm and Shattuck-Hufnagel, 2007).

The boundary marking role of phrase/utterance final glottalization has been confirmed by a number of studies. Henton and Bladon (1988) demonstrated its occurrence in sentence final position in British English RP. In American English, too, irregular vocal cord vibration is one of the acoustic cues indicating end of sentence (Slifka, 2006). In Swedish read speech, irregular phonation occurs at phrase boundaries (Fant and Kruckenberg, 1989); also in Finnish, Czech, Serbian and Croation (Lehiste, 1965). Glottalization can also signal end of turn (Redi and Shattuck-Hufnagel, 2001). In Hungarian, glottalization often occurs sentence finally, both in read and in spontaneous speech (Bőhm and Ujváry, 2008; Markó, 2009, 2010, 2011).

In English, glottalization occurs between adjacent vowels flanking a word boundary (Gimson, 1980); and word initially it often occurs before an initial vowel in English (Dilley et al., 1996). Initial vowels in German are canonically realized with a glottal stop (Rodgers, 1999), and Kohler (1994) reported a high probability of glottal onsets for vowel-initial morphemes internal to polymorphemic words as well. According to Pierrehumbert and Talkin (1992), word-initial glottalization is the most widespread if the word is itself initial in an intonational phrase.

In an earlier study on sentence-final glottalization (Markó, 2011), we have analyzed occurrences of irregular voicing in other positions. We have concluded that, in Hungarian, word initial vowels do not induce irregular voicing by themselves; but in vowel clusters across word boundaries, glottalization often crops up (although less often than in sentence final position). These results have motivated the present investigation in which vowel sequences occurring in a variety of positions (morpheme internally, across a morpheme boundary, across a compound boundary, across a word boundary) are studied, with special emphasis on the role of glottalization in separating adjacent vowels from one another.

2 Materials, methods, subjects

Subjects were asked to read 19 sentences aloud. The sentences included a total of 222 words, with 4 to 24 words per sentence. Here are a few examples: *Egyszer**ű**en nem értem, m**i**ért lenn**e** **e**z id**eá**lis* ‘I simply can’t understand why this would be ideal’. *Néh**a** **a**nnyir**a** **u**nt**a** **a** hossz**ú** **u**tazásokat, hogy elhatározta, felmond, még h**a** **e**zzel al**áá**ss**a** **i**s a további karr**ier**jét* ‘Sometimes he was so bored by the long journeys that he decided he would quit even if that would undermine his future career’. *Innen a D**eá**k térr**e** **e**gy ór**a** **a**latt s**e** **é**rsz oda* ‘You won’t make it to Deák Square from here, not even in an hour’.

The sentences included a total of 97 hiatal vowel sequences, 5.1 per sentence on average; their occurrence per sentence varied between 1 and 9. The distribution of vowel clusters by position, as well as a few examples of each type, can be found in Table 1 below.

Table 1. The vowel clusters under investigation, with examples

Position	Tokens	Vowel clusters	Examples
Morpheme internally	13	*aó, au, eá, ee, eo, ie, ió, oá, uá, üá*	*fár**aó*** [fa:rɑo:] ‘pharaoh’, *ka**lau**z* [kɑlɑuz] ‘conductor’, *id**eá**lis* [idɛa:liʃ] ‘ideal’, *t**ee**ndő* [tɛɛndø:] ‘thing to do’, *n**eo**n* [nɛon] ‘neon’, *karr**ie**r* [kɑriʲɛr] ‘career’, ***oá**zis* [oa:ziʃ] ‘oasis’, *akt**uá**lpolitikai* [ɑktualpolitikɑʲi] ‘actual political’, *n**üá**nsznyit* [nyansɲit] ‘slight-acc’
Across morpheme boundary	13	*ai, ei, ia, ié, iu, óa, ői, űe*	*kritik**ai*** [kritikɑʲi] ‘critical’, *ismeret**ei**t* [iʃmɛrɛtɛʲit] ‘knowledge-poss-acc’, *udvar**ia**s* [udvɑriʲɑʃ] ‘polite’, *Dan**ié*** [dɑniʲɛ:] ‘that of Daniel’, *szerkesztő**i**nek* [sɛrkɛstø:ʲinɛk] ‘editor-pl-dat’
Across compound boundary	31	*aa, áa, áá, ae, áo, aő, áu, ea, eá, eí, iá, ie, ió, óe, ói, óu, óú, őa, őá, őe, őú, őü, ua, úe, űe, űú*	*f**aa**rccal* [fɑɑrtsɑl] ‘with poker face’, *r**áa**dás* [ra:ɑda:ʃ] ‘encore’, *hozz**áá**llásán* [hoz:a:a:l:a:ʃa:n] ‘attitude-poss-superessive’, *haz**ae**ngedték* [hɑzɑɛŋgɛt:ɛ:k] ‘was allowed to go home’, *hozz**áo**lvas* [hoz:a:olvɑʃ] ‘read in addition’, *pály**aő**rnek* [pa:jɑø:rnɛk] ‘lineman-dat’, *r**áu**nt* [ra:unt] ‘got tired with’, *b**eí**rni* [bɛi:rni] ‘to write in’, *k**iá**llítást* [kiʲa:l:i:ta:ʃt] ‘exhibition-acc’, *úr**ie**mber* [u:riɛmbɛr] ‘gentleman’, *fal**ió**rát* [fɑlio:ra:t] ‘wall clock-acc’, *adóellenőrtől* [ɑdo:ɛl:ɛnø:rtø:l] ‘taxman-abl’, *folyó**i**rat* [fojo:irɑt] ‘periodical’, *lóugrásban* [lo:ugra:ʃbɑn] ‘knight’s move-inessive’, *hajó**ú**tról* [hɑjo:u:tro:l] ‘voyage-delative’, *fő**ú**r* [fø:u:r] ‘head waiter’
Across word boundary	40	*a#a, a#e, a#i, a#u, e#a, e#e, e#é, e#ö, e#ü, i#a, i#á, i#e, i#é, i#i, ó#a, ó#á, ó#e, ó#é, ó#o, ő#o, ú#é, ú#u, ű#i*	*ok**a a**nnak* [okɑ ɑn:ɑk] ‘reason for that’, *h**a e**zzel* [hɑ ɛz:ɛl] ‘if with this’, *annyir**a u**nta* [ɑɲ:irɑ untɑ] ‘was so bored’, *ebb**e a** könyvbe* [ɛb:ɛ ɑ køɲvbɛ] ‘into this book’, *lenn**e e**z* [lɛn:ɛ ɛz] ‘would this be’, *eszéb**e ö**tlött* [ɛse:bɛ øtløt:] ‘occurred to him’, *mellett**e ü**lő* [mɛl:ɛt:ɛ yløː] ‘sitting next to him’, *szemközt**i a**partmanban* [sɛmkøsti ɑpartmɑmbɑn] ‘in the apartment opposite’, *enged**i á**t* [ɛŋgɛdi a:t] ‘surrenders it’, *szeptember**i e**lőadáson* [sɛptembɛri ɛløːɑda:ʃon] ‘at the September performance’, *korább**i i**smereteit* [kora:bːi iʃmɛrɛtɛʲit] ‘his previous knowledge-acc’, *oktat**ó á**llandóan* [oktɑto: a:l:ɑndo:ɑn] ‘instructor permanently’, *helyreállít**ó o**peráció* [hɛjrɛa:l:i:to: opɛra:tsiʲo:] ‘reconstructive operation’

The sentences were read aloud by university students and by adults with a university degree. The ages of the 10 subjects were between 21 and 51 years (their mean age was 30 years) and an equal number of male and female subjects were involved. Their articulation and hearing were normal and their voice production did not show any pathological traits (according to their self-report and the experimenter’s informal observation). Their reading performance was recorded under laboratory conditions. The subjects had a chance to read the sentences for

themselves first, and the recording started when they indicated they were ready for it. Yet, uncertainties did occur in their production, including uncorrected misreading and slips of the tongue. Thus, 95 or 96 vowel sequences were correctly read aloud per subject, and eventually a total of 959 occurrences were analyzed. The recordings were labeled with the help of the Praat program (version 5.1; Boersma and Weenink, 2009). The vowel sequences were analyzed according to the following criteria:

(1) the position of the cluster: morpheme internal, across morpheme boundary, across compond boundary, across word boundary;

(2) the phonetic quality of the vowels constituting the cluster;

(3) the realization of the boundary between the vowels: plain transition, hiatus resolution, deletion of one of the vowels, glottalization, pause, and any combination thereof (e.g. hiatus resolution + glottalization, pause + glottalization).

The analysis of glottalized realizations was performed in accordance with the methodology of previous studies (e.g. Dilley et al., 1996; Bőhm and Ujváry, 2008), combining visual and auditive information. In Praat, the waveform and the spectrogram were displayed, as were pitch and intensity curves where necessary (together with the labels, of course); the sound recordings were continuously listened to and double checked. Acoustically, a given speech sound segment was taken to be glottalized if

(1) the duration or amplitude of basic periods suddenly changed to a significant extent (including the occurrence of a glottal stop, Dilley et al., 1996); or if

(2) the fundamental frequency suddenly fell below the speaker's normal/usual pitch range (according to the author's auditory judgement).

Figure 1 shows realizations of the same vowel cluster from two male subjects' production, in which the difference between the modal voiced and the glottalized transition can be detected.

In addition, as a perceptual criterion, we took into consideration cases in which the timbre of the segment was audibly hoarse or creaky. We labeled a segment 'glottalized' when its irregularity was both visible in the displays and audible in the sound recording.

Recall (from the Introduction) that glottalization is a multifunctional phenomenon. Thus, the fact that a given vowel sequence was realized with irregular voicing does not necessarily mean that glottalization had the function of boundary marking between the two vowels. In order to avoid misinterpretation of the data, occurrences in which the whole V(#)V sequence was glottalized, or in which one of the vowels and its environment (the preceding/following syllable(s)) were realized with irregular phonation, were not taken into account as glottalized – this involved 11.2% of all V(#)V sequences realized in the material. These realizations were categorized as plain transition, hiatus resolution, etc., as appropriate. Note that this level of occurrence corresponds to the average frequency of glottalization in any Hungarian speech sample (Bőhm and Ujváry, 2008; Markó, 2011) and that such realizations primarily occurred as the speakers got closer to the ends of sentences.

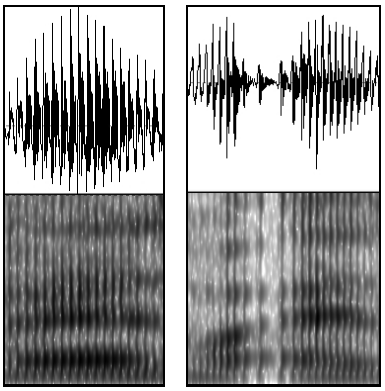


Figure 1. Two occurrences of A#E sequence in *ha ezzel* (‘if with this’): modal voiced (left) and glottalized (right) realizations

In sum, we took the following types of realizations into account as instances of glottalization (presumably intended to demarcate vowels in hitaus):

– the end of V_1 (less frequently, the whole of V_1) was glottalized and V_2 had modal voice;

– V_1 had modal voice and the beginning of V_2 (less frequently, the whole of V_2) was glottalized;

– the end of V_1 and the beginning of V_2 were both glottalized (including cases where V_2 started with a glottal stop, after a brief pause).

We analyzed the effects of individual pronunciation habits, too. For the statistical analyses (correlation analysis), we used version 15.0 of SPSS.

3 Results

3.1 The position of the vowel sequence and the quality of the vowels involved

As expected, most vowel sequences were realized with plain transitions both morpheme internally and across compound/word boundaries (Figure 2). Across morpheme boundaries, hiatus resolution was the most frequently occurring realization type, due to the fact that V + *i* and *i* + V sequences are fairly frequent in this position in everyday Hungarian speech and, accordingly, also in the present corpus. In Szende (1973)’s data, the most frequent (word internal) vowel sequences are *ia, ai, iá, ie, ió, ei;* only the seventh most frequent combination is one without an *i: óa*.

Realizations were the most varied across word boundaries, since in this position pauses (and pauses with glottalization) were also an option. Glottalization was also the most frequent in this position; but its occurrences were found morpheme internally and across compound boundaries as well.

Let us turn to a more detailed investigation of the V(#)V sequences occurring at the various levels, including the quality of the vowels involved and the type of realization attested.

3.1.1 Morpheme internal vowel sequences

Morpheme internally we analyzed a total of 130 vowel clusters. Of these, 66.9% (87 tokens) were realized with a plain transition, 21.5% (28 tokens) with hiatus resolution with *j*, and 6.9% (9 tokens) with glottalization. In 4.6% of the clusters (6 tokens), one of the vowels was deleted.

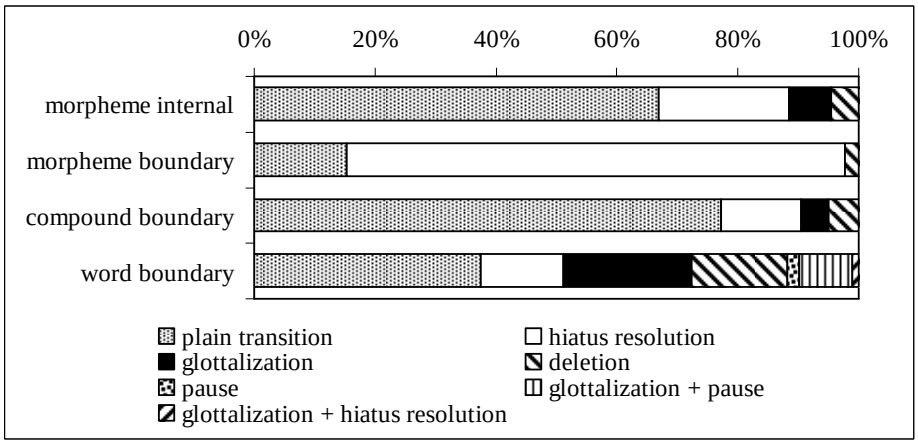


Figure 2. The distribution of realization types in V(#)V sequences of various levels

Obviously, vowel quality unambiguously affected the occurrence of hiatus resolution (Figure 3): clusters involving *i* were invariably realized in this manner, as expected (*operác**ió*** ‘operation’, *karr**ie**r* ‘career’). In defiance of the phonological regularity, however, we found hiatus resolution in 26.7% of *eá* sequences, too: 6 speakers pronounced the name *D**eá**k* in this manner, and 1 speaker (each) pronounced *id**eá**lis* ‘ideal’ and *óc**eá**ni* ‘ocean-adj.’ with hiatus resolution.

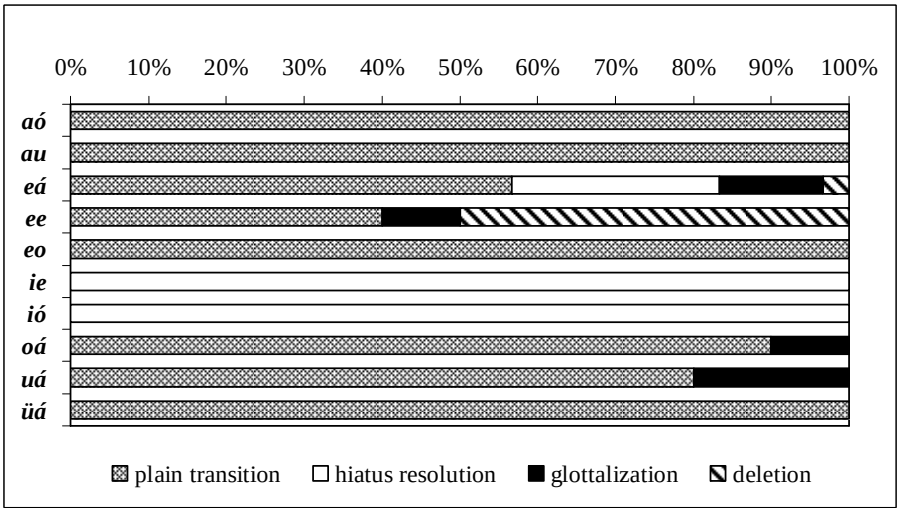


Figure 3. Types of realizations of morpheme-internal vowel sequences

The sequences *aó, au, eo, üá* were invariably realized with plain transitions (e.g. *fár**aó*** ‘pharaoh’, *kal**au**z* ‘conductor’, *n**eo**n* ‘neon’, *n**üá**nsz* ‘nuance’). The fact that half of the speakers pronounced the word *t**ee**ndőkről* ‘of things to do’ with a single long [ɛː] was counted as an instance of deletion. (On the other hand, cases in which the speaker demarcated the vowels by modulating the fundamental frequency – and which were therefore perceived as including two distinct vowels – were counted as sequences of two vowels even if the waveform or the spectrogram did not exhibit any change at the critical point). In addition, one speaker realized *óc**eá**ni* ‘ocean-adj.’ as [oːtsaːni], without [ɛ].

We found glottalization in four types of clusters: *eá, ee, oá, uá*. Two speakers (a different two in each case) used irregular phonation to demarcate the vowels in *id**eá**lis* ‘ideal’, *ob**oá**s* ‘oboist’, and *akt**uá**lpolitikai* ‘actual political’. We found a single glottalized realization for *óc**eá**ni* ‘ocean-adj.’, *D**eá**k* (name) and *t**ee**ndőkről* ‘of things to do’ (again with different speakers in each case). On the basis of these data (there being too few), we do not see any tendency in which pairs of vowel qualities are (more) characterized by glottalized transitions; these data may well reflect individual pronunciation habits.

3.1.2 Vowel clusters across morpheme boundaries

In this position, again, all 130 vowel sequences were realized. Due to the large number of *i*-clusters, the proportion of hiatus resolution was exceedingly high: 82.3% (107 tokens). Another 15.4% of the sequences (20 tokens) were realized with plain transitions, and one of the vowels was deleted in 2.3% of the cases (3 tokens). In this category, glottalization did not occur at all.

The sequences *ai, ei, ia, őa* were broken up by *j* in all cases (e.g. *kritik**ai*** ‘critical’, *nőügy**ei*** ‘his affairs with women’, *udvar**ia**s* ‘polite’, *szerkesztő**i**nek* ‘his editors-dat’). In the cases of *ié* and *iu*, we also found deletions (Figure 4): *m**ié**rt* [meːrt] ‘why’ and *miután* [mutaːn] ‘since’.

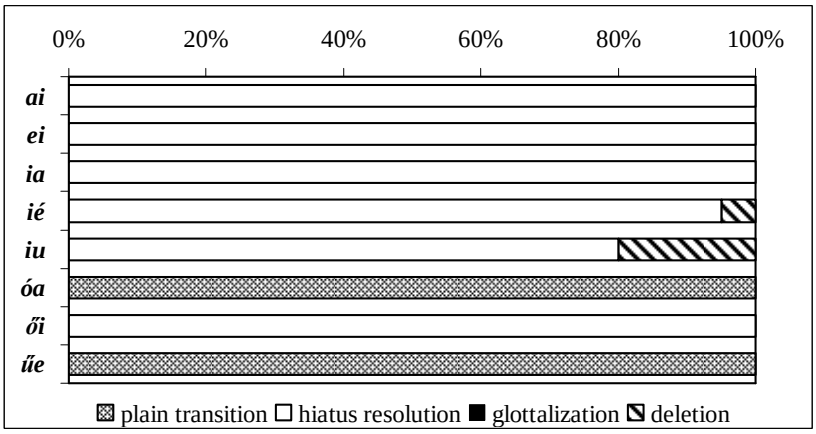


Figure 4. Types of realizations of vowel sequences across morpheme boundaries

Clusters not involving an *i* (*állandó**a**n* ‘permanently’, *egyszer**ű**en* ‘simply’) were produced with plain transitions in all speakers’ productions (despite the fact that, in spontaneous everyday speech, realizations of such items with hiatus resolution can be observed with increasing frequency).

3.1.3 Vowel clusters across compound boundaries

We analyzed 304 VV sequences across compound boundaries, 77.3% (235 tokens) of which involved plain transitions and 13.2% (40 tokens) involved hiatus resolution. The remaining nearly 10% were divided between deletion (7.9%, 24 tokens) and glottalization (4.6%, 14 tokens).

Plain transitions occurred in all vowel sequences but one (Figure 5) and were exclusively contained within *ae, áo, aő, áu, őa, őú, őü, ua, úe* (e.g. *haz**ae**ngedték* ‘they let him go home’, *hozz**áo**lvas* ‘read in addition’, *pály**aő**rnek* ‘lineman-dat’, *r**áu**nt* ‘got tired of it’, *el**őa**d(ni/áson)* ‘to perform/at a performance’, *f**őú**r* ‘head waiter’, *n**őü**gyeit* ‘his affairs-acc with women’, *kap**ua**ljban* ‘in the gateway’, *búcs**úe**stre* ‘to a farewell party’).

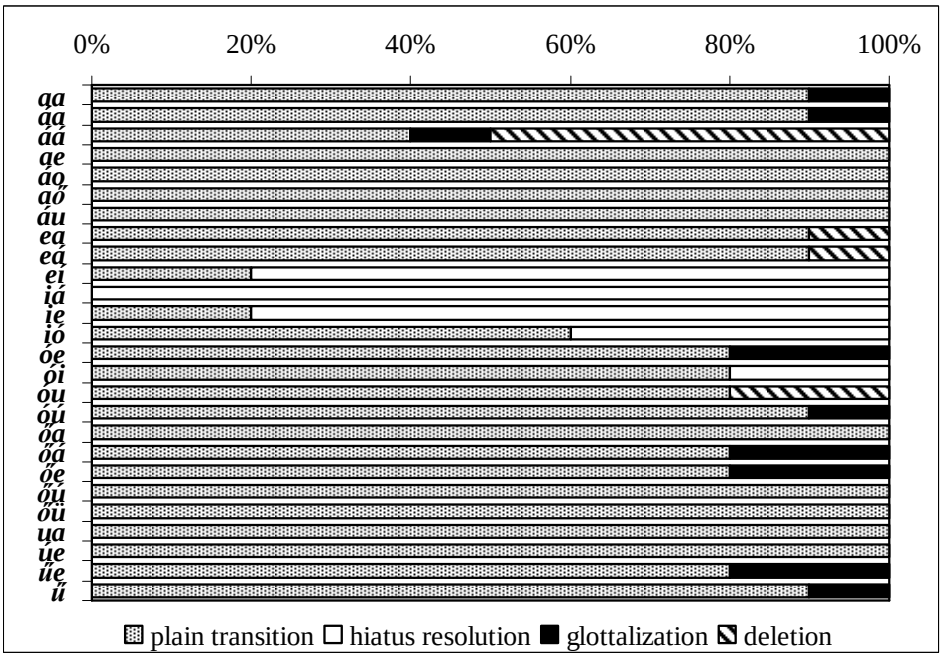


Figure 5. Types of realizations of vowel sequences across compound boundaries

Unlike in the former cases involving clusters containing *i/í*, it was only *iá* (as in *k**iá**llításon* ‘at an exhibition’) that was realized with hiatus resolution in all subjects’ speech. In the clusters *eí, ie, ió, ói*, the occurrence of *j* was variable. The word *b**eí**rni* ‘to write in’ was pronounced with hiatus resolution by 8 speakers. The words *úr**ie**mber* ‘gentleman’ and *k**ie**gészíti* ‘complements it’ were pronounced with a *j* by 9 and 7 speakers, respectively, hence the cluster *ie* also exhibited hiatus resolution in

80% of all cases. Clusters in which *i* was accompanied by *ó* were realized with hiatus resolution rather less frequently (*fal**ió**rát* ‘wall clock-acc’ in 4, and *foly**ói**rat* ‘journal’ in 2 speakers’ performance). These findings suggest that the phenomenon of hiatus resolution – as opposed to what the general phonological rule claims – is not totally independent of the quality of the other vowel of the cluster or of the type (or lack) of morpheme boundary occurring between the two vowels.

In this group, deletion was the most frequently attested for the cluster *áá* (50%), but with a significant difference between the two words concerned. The word *hozz**áá**llásán* ‘on his attitude’ was pronounced with a single [a:] by 8 speakers, whereas *al**áá**ssa* ‘he undermines it’ by only 2 speakers.

The *u* of the VV cluster of *l**óu**grásban* ‘in a knight’s move’ was deleted in the pronunciation of 2 speakers, while it was reduced to a glide by several subjects. Since such diphthongization was only observable in this one word, we did not open a separate category for it – we counted these instances among cases of plain transition.

In the cases of *ea* and *eá*, we can take the deletion of *e* to be individual speech habits as these were observable with the same speaker, in the words *bel**ea**d(ta/ott)* ‘he put (it) in’ and *helyr**eá**llító* ‘reconstructive’.

Glottalization was found in 20% of the occurrences of 4 clusters (*óe, őá, őe, űe*) and in 10% of those of 5 others (*aa, áa, áá, óu, űú*). Of the latter cases, four occurred in the production of the same speaker. The following words exhibited glottalization in two speakers’ rendering each: *al**áá**ssa* ‘he undermines it’, *ad**óe**llenőrtől* ‘from the tax inspector’, *el**őá**llt* ‘presented itself’, *es**őe**mberre* ‘to the rain man’, *bet**űe**jtés* ‘spelling pronunciation’.

3.1.4 Vowel clusters across word boundaries

In this position, 395 V#V clusters were analyzed. Again, the most frequent solution was plain transition, but far less frequently than in the case of morpheme internal clusters or those across a compound boundary. As compared to the 37.5% (148 tokens) share of plain transitions, glottalized realizations were also frequent, 31.1% (123 tokens) taken together. Within that percentage, VV clusters involving glottalization alone represented 21.3% (84 tokens), glottalization was combined with a pause in 8.9% of the cases (35 tokens), and it even occurred with hiatus resolution, in 1.0% (4 tokens). Deletion was documented in 15.7% of the cases (62 tokens), hiatus resolution in 13.7% (54 tokens), and pause (by itself) in 2.0% (8 tokens).

It was only in the case of *e#ü* (*mellett**e ü**lő* ‘sitting next to him’) that plain transition occurred in all speakers’ production (Figure 6), though 90% frequency was found for *e#é* (*s**e é**rsz* ‘nor do you arrive’). Apart from a few items involving *i*, it was only in the case of *ó#á* (*oktat**ó á**llandóan* ‘instructor permanently’) that plain transition did not occur at all.

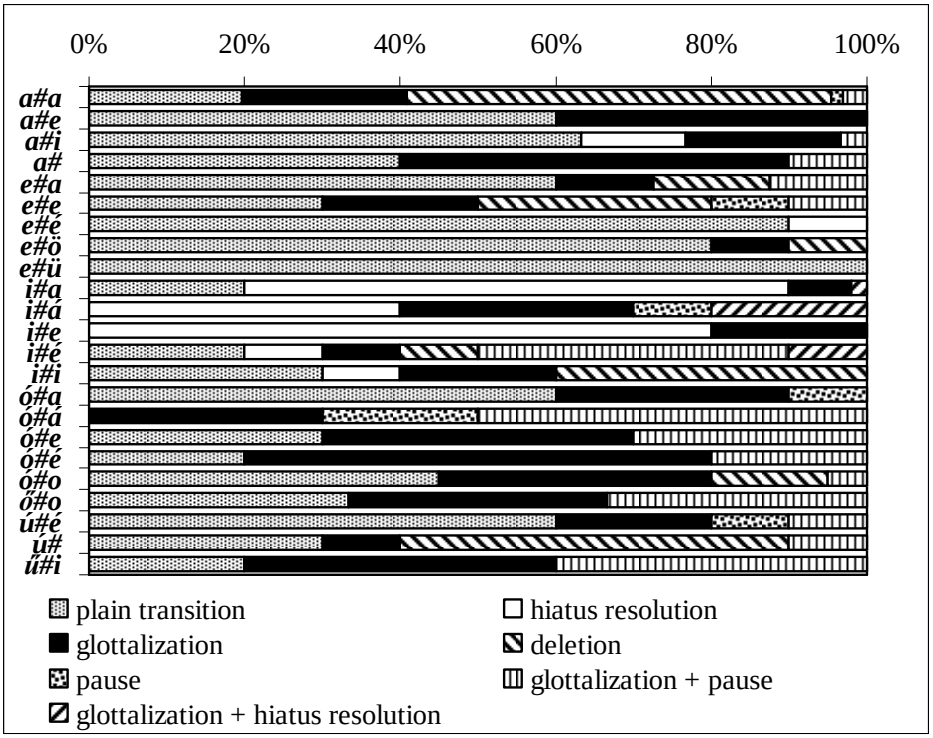


Figure 6. Types of realizations of vowel sequences across word boundaries

Hiatus resolution by *j* occurred in this position far less frequently than in the word internal positions, even in cases where one of the vowels was *i*. We did not find 100% hiatus resolution in any of the cluster types. Although almost all instances of the cluster *i#a* were pronounced with hiatus resolution by almost all subjects (*mi* ***a****z* 'what is it' 9[1], *aki* ***a*** 'who the' 8, *intézi* ***a*** 'organizes the' 8, *kiegészíti* ***a*** 'complements the' 10), there was an instance of this cluster that never contained a transitional *j* (*szemközti* ***a****partmanban* 'in the apartment across the corridor'). The second vowel in sequences often pronounced with hiatus resolution was the definite article in most cases; it can be assumed that this is a motivating factor for hiatus resolution across a word boundary. The same strategy was followed in 80% of the instances of *i#e* (*szeptember* ***e****lőadáson* 'at a September performance'), whereas only 4 speakers pronounced *i#á* (*enged* ***á****t* 'allows him to go through') with a *j*. The sequence *a#i/í* was performed with hiatus resolution in just one example out of three, in a few speakers' pronunciation (*alá* ***i****s* 'undermines, too' 4, but *haz* ***i****dőben* 'home in time' and *belg* ***í****ró* 'Belgian writer' both 0). A single speaker (each) pronounced *j* in *művészet* ***é****s* 'artistic and', *korább* ***i****smereteit* 'his earlier knowledge'; no one did so in *cím* ***i****rodalmi* 'literary (journal) entitled'. Of all clusters involving *é,* only *s* ***é****rsz*

[1] The numbers refer to how many subjects produced the given realization.

‘nor do you arrive’ was realized with hiatus resolution, and only in one case. Hiatus resolution was combined with glottalization in 4 *i#*V sequences.

It was in this position that deletion occurred the most often, probably due to the facts that this strategy is typical with respect to clusters of identical vowels and that the appropriate input configuration occurs mostly across word boundaries. The sequence *aa* was coalesced at least once in each example, in a total of 54.6% of all tokens (in *ór**a a**latt* ‘in an hour’ by 8 speakers, in *unt**a a*** ‘was bored by the’ by 6, *list**a a*** ‘list the’ by 6, *arr**a a**z* ‘to it the’ by 6, *beleadt**a a**z* ‘put in the’ by 5, *ok**a a**nnak* ‘reason for that’ by 4, and *néh**a a**nnyira* ‘sometimes so much’ by 1 speaker). The sequence *ú#u* (*hossz**ú u**tazásokat* ‘long journeys-acc’) was realized with a single vowel in 50% of the cases, *i#i* (*korább**i i**smereteit*) ‘his earlier knowledge-acc’ in 40%, *e#e* (*lenn**e e**z* ‘would this be’ with 5 speakers, *térr**e e**gy* ‘to the square in a’ with 1 speaker) in 30%, and *ó#o* (*helyreállít**ó o**peráció* ‘reconstructive operation’ with 2 speakers, *jó **o**ktató* ‘good instructor’ with 1 speaker) in 15% of the cases. In clusters of dissimilar vowels, the first vowel was deleted in 15% in *e#a* (in *ebb**e a*** ‘into this’ 5 times, in *rendezőj**e a**z* ‘its director the’ once, in *esőemberr**e a*** ‘rain man the’ and in *b**e a*** ’into the’ not at all); and in 10% of *i#é* (*művészet**i é**s* ‘artistic and’ with a single speaker).

Pauses occurred by themselves in 2.0% of the cases (8 tokens), and accompanied by glottalization in 8.9% (35 tokens), so that, word boundaries involving pauses amounted to over 10% of all vowel clusters flanking a word boundary. The duration values of the two types of pauses were also quite dissimilar. The average duration of plain pauses (those not involving glottalization) was 118 ms (SD 79 ms); that of pauses with glottalization was 68 ms (SD 52 ms). Thus, pause duration and the function of irregular phonation appear to be interrelated, in that the latter reinforces the demarcation effect of short pauses. It follows that glottalized V#V sequences where an extremely long pause occurred between the two vowels are to be treated separately in further analysis as in these cases the function of glottalization is different from the demarcating and disambiguating function that we can assume for the other instances.

Among word combinations separated exclusively by a pause (*oktat**ó*** | ***á**llandóan* ‘instructor permanently’ and *térr**e*** | ***e**gy* ‘to the square in a’ occurred with two speakers); and each of the following was pronounced in that manner by a single speaker: *unt**a*** | ***a*** ‘was bored by the’; *enged**i*** | ***á**t* ‘allows to pass’; *sz**ó**,* | ***a**mi* ‘word that’; *hossz**ú*** | ***é**lménybeszámolót* ‘lengthy account-acc’. There seems to be no clear tendency in terms of vowel quality, and although the marking of phrase boundaries may influence the occurrence of pauses somewhat, this is not an unambiguous motivational factor, either.

Glottalization occurred with very high frequency in vowel clusters spanning a word boundary. This occurred in nearly one third of the relevant data where we tested it either alone or combined with a pause or with a transitional *j*. Phrase boundaries were positions in which glottalized word boundaries occurred in the largest proportion, with 8 speakers in each of the following examples: *a jó oktat**ó***

***á**llandóan hozzáolvas* ‘a good instructor permanently reads in addition’; *a helyreállító operáci**ó** **é**rtelmetlen lesz* ‘the reconstructive operation will be pointless’; *a Műút cím**ű** **i**rodalmi, művészeti és kritikai folyóirat* ‘the literary, artistic and critical journal entitled Highway’. The following items, likewise including a phrase boundary, were realized with irregular phonation in 70% of the cases: *a belga író **e**bbe a könyvbe* ‘the Belgian writer into this book’; 66.7%: *az utcán lévő, **o**ázist ábrázoló neonok* ‘the neon signs in this street, depicting an oasis’; 60%: *a Deák térr**e** **e**gy óra alatt* ‘to Deák Square, within an hour’; *művészet**i** **é**s kritikai* ‘artistic and critical’. In the cases of *néh**a** **a**nnyira* ‘sometimes so much’ and *annyir**a** **u**nta* ‘was so bored’, glottalized implementations (that also occurred 60% of the time) may be due to stress (see below). Among the cases listed here, instances in which glottalization was combined with a pause were also found in a varying degree (1 to 5 speakers).

In this position, there were only two word combinations that none of the subjects realized with glottalization: *s**e** **é**rsz* ‘nor do you arrive’ and *mellett**e** **ü**lő* ‘sitting next to her’. In these cases, the reason why boundary marking was not necessary may have been the clitic-like relationship between the words.

On the basis of the foregoing results, it appears that the grammatical (and hence the prosodic) structure of the construction must have influenced the implementation of the vowel sequences spanning a word boundary. Therefore, we went back to the other cases and looked at them again in light of this factor. Phrase boundaries led to irregular phonation in quite a few cases: at least half of the speakers glottalized V#V clusters within which they detected an intonational phrase boundary. Note that it was not the case that all speakers rendered the sentences with the same intonational structure. For instance, the sentence *A belga író ebbe a könyvbe beleadott apait-anyait* ‘The Belgian writer wrote this book with might and main’ was pronounced as two intonational phrases by some subjects and as three by others. Due to its strong emotional accents, the clause *Néha annyira unta a hosszú utazásokat…* ‘He was sometimes so bored by his long journeys…’ was broken up into as many as three or four intonational phrases. On the other hand, some clauses failed to form an independent intonational phrase of their own. For instance, speakers (with one exception) did not stop when seeing the first comma in *Az első szó, ami Kádárról eszébe ötlött, a „betűejtés” volt* ‘The first word that came to his mind concerning Kádár was “spelling pronunciation”’, and only three speakers marked the clause boundary by glottalization.

In some cases, glottalization was noted even inside clitic groups, despite their prosodic unity. In *enged**i** **á**t* ‘lets him pass’ and *belg**a** **í**ró* ‘Belgian writer’, 5 and 4 speakers, respectively, changed their voice quality, and in *szemközt**i** **a**partmanban* ‘in the apartment across the corridor’ and *haz**a** **i**dőben* ‘home in time’, 3 speakers did so. This was probably in order to avoid realizations involving hiatus resolution (that overwhelmingly characterized the rest of the clusters involving *i*, even across a word boundary, see above). In other cases, the speakers probably wanted to avoid vowel deletion. This is the most likely explanation for clusters of identical vowels

(and of vowels merely differing in length): *jó **o**ktató* ‘good instructor’, *helyreállító **o**peráció* ‘reconstructive operation’, *ok**a** **a**nnak* ‘reason for that’ (4 glottalized realizations in each item). We saw examples of deletion in *e#a* clusters, too, so the irregular phonation occurring at the boundary between these two low vowels may also be a deletion-avoiding strategy, as in *h**a** **e**zzel* ‘if with this’ (4 speakers). These cases suggest an attempt at careful pronunciation on the speakers’ part.

3.2 Forms of glottalization

We found a total of 150 glottalized V(#)V sequences. In those combined with a pause 200 ms or longer, the long pause was in itself capable of fulfilling a demarcation function; hence, these tokens were excluded from further analysis. In all four cases of this type, the vowel started with a glottal stop after the pause.

Of the 146 remaining vowel sequences, 9 were realized morpheme internally, 15 at a compound boundary, and 122 on both sides of a word boundary. The distribution of these items across realization types is shown by Figure 7, where implementations involving a brief pause are also included.

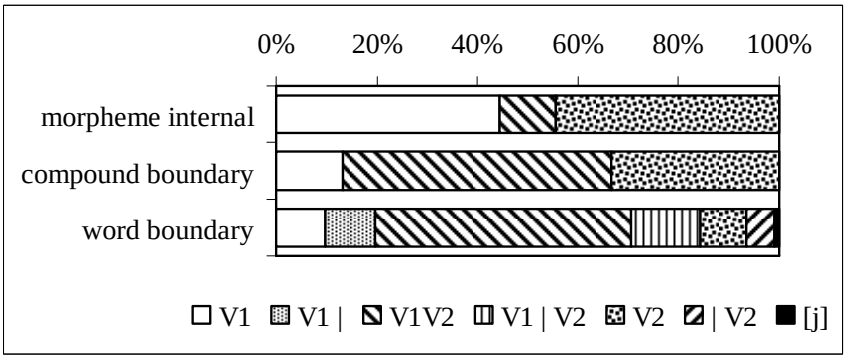


Figure 7. Types of realizations of glottalization at the various levels (V1 = only the end of V1 glottalized; V1| = only the end of V1 glottalized, followed by pause; V1V2 = the vowel transition glottalized; V1|V2 = the transition glottalized, with an intervening pause; V2 = only the beginning of V2 glottalized; |V2 = only the beginning of V2 glottalized, preceded by pause; [j] = the hiatus filler glottalized)

In the case of VV clusters realized with irregular phonation morpheme internally, usually only one of the vowels was (wholly or partially) glottalized (in 4 cases each, amounting to 44.4% of the whole), and only one case (11.1%) was found where it was the transition between the two vowels that was realized with irregular voicing. Across compound boundaries, a glottalized transition was the more typical solution (53.3%), followed by cases in which the beginning (less often, the whole) of the second vowel was irregular (33.3%). Cases where only the end of the first vowel was glottalized amounted to 13.3%.

Glottalization combined with a pause only occurred across word boundaries, of course. Also, we found a single realization in this position (0.8%) in which the

transitional *j* itself was glottalized (*engedi*[j]*át* 'allows to pass'). In roughly half of the cases, irregular phonation occurred at the boundary between the two vowels again (50.8%). If we add occurrences in which the vowels were separated by a pause (while the end of V_1 and the beginning of V_2 were both glottalized), this percentage rises to 64.8%. Only the first vowel was glottalized (partially, less often wholly) in 19.7% of the cases. Within this group, there were equal numbers of cases in which V_1 was or was not followed by a pause. In 14.8% of the cases, only the beginning of V_2 (or, less often, the whole of it) was glottalized; where 9.0% were realized without a pause, and 5.7% with a pause before the second vowel.

We have also examined in what parameters irregularity of phonation revealed itself. In Figure 8 we show a few concrete realizations. In part *a)* of the figure, we see periods of variable amplitude following one another unsystematically at the boundary of the two identical vowels. In part *b),* the periods at the end of V_1 become few and far between, then voicing ceases while the configuration of the vocal tract remains stable for a while; then, following a pause, the speaker articulates the second vowel. Part *c)* of the figure shows diplophony.

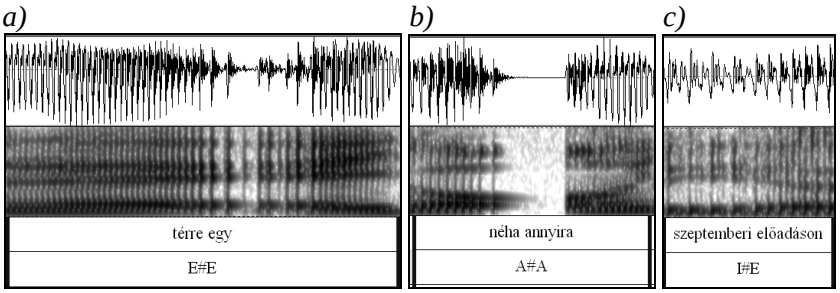


Figure 8. Examples of irregular phonation

3.3 Variability across speakers

As we have seen, in 11.2% of V(#)V clusters irregular phonation occurred without any intention to demarcate vowels from one another but presumably motivated by some other communicative intention or just some individual pronunciation habit of the speaker. This claim can be based on the fact that, in such cases, either the whole vowel sequence was glottalized throughout (hence irregular phonation could not have a separating function in these cases) or irregular phonation was spread out to adjacent speech sounds or even sequences of several syllables. The latter cases mainly occurred at the ends of sentences or phrases.

Frequency of glottalization as characteristic of the individual speakers has been studied by looking at such V(#)V sequences and the proportions of occurrence of the two functions of irregular phonation speaker by speaker. The frequency of occurrence of irregular phonation in a demarcating function (labeled as 'glottalization' in Figure 9) and in some other role ('not analyzable' in the figure) differed significantly across speakers. Half of the subjects (M2, M5, F1, F2, F3)

used glottalization for demarcative purposes much more often (at least twice as frequently) than for some other reason. An extreme value was exhibited in this respect by F2 who used irregular voicing 30 times as often for separating adjacent vowels than for any other purpose.

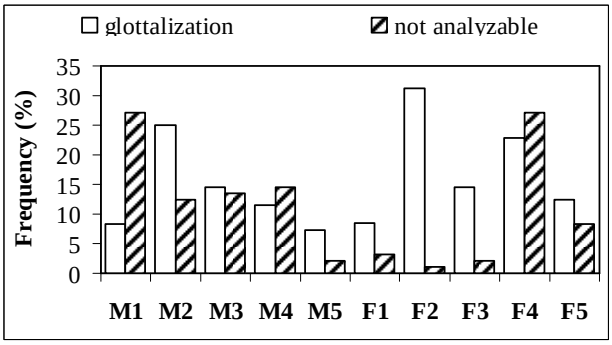


Figure 9. The frequency of glottalization with individual speakers (M = male, F = female)

Four speakers (M3, M4, F4, F5) used glottalization to a similar extent in both functions. Although the voice production of M1 was characterized by irregularity to a large extent, he only separated vowels by this strategy in one fourth of his glottalized V(#)V sequences; in his case, irregular voice production was rather a kind of individual speech habit.

We also compared the distribution of the realizations of V(#)V sequences across speakers (Figure 10). Of course, all speakers' productions were dominated by plain transitions, in 34.4 to 54.2%. Hiatus resolution, as expected, was present in nearly identical proportions, given the fact that it is regulated by a phonological rule; in addition, it was also determined to some extent by individual speech habits (20.8–28.1%). Inter-speaker variance was observable to the largest extent with respect to glottalization, occurring in 7.3–31.3% of the cases. Similarly large differences were found in the frequency of deletion: 3.1–19.8%. Pauses were relatively rare in all speakers' productions (0.0–9.4%), partly due to the fact that most VV clusters were word internal in our corpus.

We saw that glottalization worked as a strategy for avoiding hiatus resolution or deletion in some cases; therefore, we tried to find out if there was any correlation between this phenomenon and other types of realization in their frequency across speakers. A Pearson's correlation yielded a significant, strong, negative correlation ($r = -0.708$, $p = 0.022$) between glottalization and deletion, supporting this hypothesis. In the case of hiatus resolution, we did not find any correlation (restricting our data set to clusters involving *i*).

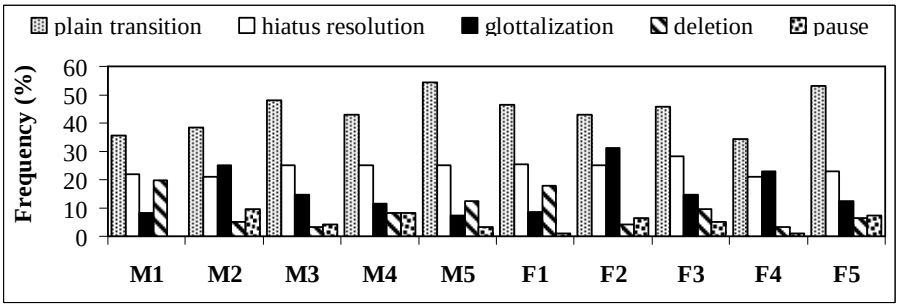


Figure 10. Occurrence of types of V(#)V sequences across speakers (M = male, F = female)

4 Conclusions

In this paper, we studied the realization of vowel sequences in Hungarian speech, with special regard to the role of glottalization in separating adjacent vowels from one another. In such cases, hiatus was not resolved by the speaker but rather enhanced for some reason (like indicating a phrase boundary).

We studied the interdependence of the type of realization and the position of the VV cluster, as well as that between the former and the quality of the vowels involved. We analyzed the forms of glottalization and the effect of individual pronunciation habits.

Glottalization occurred the most frequently across word boundaries; sometimes (with a frequency below 10%). It also happened morpheme internally or across compound boundaries. Glottalized word transitions were realized mostly at phrase boundaries (stress also influenced the occurrence of glottalization). Another major motivation for a glottalized realization of V(#)V clusters was to avoid the use of some phonological/articulatory mechanism (hiatus resolution or deletion). In the case of deletion, this was also statistically confirmed.

In a significant majority of cases, glottalization occurred in the transitional region between two vowels and was mainly implemented as temporal variability in the periods of vocal cord vibration.

A large amount of inter-speaker variability was shown in the frequency of occurrence of glottalization in its diverse functions. The question arises whether speakers employ this type of boundary marking intentionally. In view of the results, we can assume some kind of intentionality with respect to demarcation (i.e., where exactly the speaker wishes to insert some kind of boundary marking), but for the choice of the manner of boundary marking (whether the speaker opts for glottalization or some other strategy at any given moment), it would be hard to assume any kind of intentionality. The occurrence of glottalization in these positions is probably due to the fact that the speaker wishes to suspend voice production at the given point in order to properly separate adjacent segments from one another – but this would take up too much time and energy. So the movement of the vocal cords does not exactly reach the configuration of a glottal stop, it only comes close to that

state. This explanation is supported by Peter Ladefoged's theory that phonation types can be ordered along a continuum without discrete cut-off points (Ladefoged 1971). The two extreme ends of that continuum would be 'fully open glottis' and 'fully closed glottis'. The former yields voiceless consonants, and as the vocal cords come increasingly closer to one another, we get – via breathy, modal, and creaky voice – to the other end: glottal closure (Figure 11).

Our results show that glottalization occurs relatively frequently in Hungarian speech, and therefore it influences the data/results of both segmental and suprasegmental phonetic analyses. It has a possible demarcating function in the case of V#V clusters just as much as in indicating sentence/utterance ends.

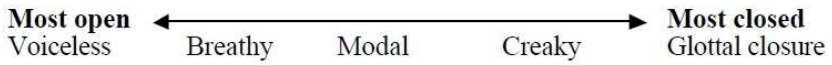


Figure 11. A continuum of phonation types (Gordon and Ladefoged 2001: 384)

In this paper, we analyzed vowel sequences realized in read sentences. The ultimate aim of carrying on with our studies of this phenomenon in Hungarian speech is to extend our knowledge of the communicative functions of voice quality in general.

5 Acknowledgements

I would like to thank Péter Siptár for the English translation and for the numerous pieces of advice I received from him. Work on this paper was supported by Bolyai Research Grant #BO/00093/09.

References

Batliner, A., Burger, S., Johne, B. and Kiessling, A. 1993. MÜSLI: A classification scheme for laryngealizations. In: *Working Papers, Prosody Workshop*. Sweden: Lund. 176–179.

Batliner, A., Steidl, S. and Nöth, E. 2007. Laryngealizations and emotions: How many Babushkas? In Schröder, M., Batliner, A. and d'Alessandro, Ch. (eds.): *Proceedings of the International Workshop on Paralinguistic Speech (ParaLing'07, Saarbrücken 03.08.2007).* Saarbrücken: DFKI. 17–22. http://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2007/Batliner07-LAE.pdf

Boersma, P. and Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.2). http://www.fon.hum.uva.nl/praat/download_win.html

Bőhm, T. and Shattuck-Hufnagel, S. 2007. Listeners recognize speakers' habitual utterance final voice quality. In Schröder, M., Batliner, A. and d'Alessandro, Ch. (eds.): *Proceedings of the International Workshop on Paralinguistic Speech (ParaLing'07, Saarbrücken 03.08.2007).* Saarbrücken, 29–34. http://www.bohm.hu/publications/Bohm-ShattuckHufnagelParaling2007.pdf

Bőhm, T. and Ujváry, I. 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben [Irregular phonation as an individual speaker's characterisic in Hungarian speech]. *Beszédkutatás 2008.* 108–120.

Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–444.

Fant, G. and Kruckenberg, A. 1989. Preliminaries to the study of Swedish prose reading and

reading style. *Speech Transmission Laboratory Quarterly Progress and Status Report* 30/2. Stockholm: Royal Institute of Technology. 1–80. http://www.speech.kth.se/prod/publications/files/qpsr/1989/1989_30_2_001-080.pdf
Gimson, A. Ch. 1980. *An introduction to the pronunciation of English.* 3rd edition. London: Edward Arnold.
Gobl, Ch. and Ní Chasaide, A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication 40,* 189–212.
Gordon, M. and Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383–406.
Gósy, M. 2004. *Fonetika, a beszéd tudománya* [Phonetics: The science of speech]. Budapest: Osiris Kiadó.
Grivičić, T. and Nilep, Ch. 2004. When phonation matters: The use and function of *yeah* and creaky voice. *Colorado Research in Linguistics* 17/1, 1–11. http://www.colorado.edu/ling/CRIL/Volume17_Issue1/paper_GRIVICIC_NILEP.pdf
Henton, C. and Bladon, A. 1988. Creak as a sociophonetic marker. In Hyman, L. M. and Li, Ch. N. (eds.): *Language, speech and mind. Studies in honour of Victoria A. Fromkin.* London and New York: Routledge. 3–29.
Hopper, P. J. and Traugott, E. C. 2003. *Grammaticalization.* Cambridge: Cambridge University Press.
Kohler, K. J. 1994. Glottal stops and glottalization in German. *Phonetica* 51, 38–51.
Ladefoged, P. 1971. *Preliminaries to linguistic phonetics.* Chicago: University of Chicago.
Lehiste, I. 1965. Juncture. In: *Proceedings of the 5th International Congress of Phonetic Sciences, Münster 1964.* New York: S. Karger. 172–200.
Markó, A. 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége* [Some suprasegmental characteristics of spontaneous speech]. PhD thesis. Budapest: ELTE.
Markó, A. 2009. Stigmatizált hanglejtésforma a spontán beszédben [A stigmatized intonation contour in spontaneous speech]. *Beszédkutatás 2009.* 88–106.
Markó, A. 2010. A prozódia szerepe a spontán beszéd tagolásában [The role of prosody in the organization of spontaneous speech]. *Beszédkutatás 2010.* 82–99.
Markó, A. 2011. A glottalizáció határjelző szerepe a felolvasásban [Boundary marking by glottalization in reading aloud]. Beszédkutatás 2011. 31–45.
Pierrehumbert, J. and Talkin, D. 1992. Lenition of /h/ and glottal stop. In Doherty, G. J. and Ladd, D. R. (eds.): Papers in laboratory phonology II: Gesture, segment, prosody. Cambridge: Cambridge University Press. 90–117.
Redi, L. and Shattuck-Hufnagel, S. 2001. Variation in the realization of glottalization in normal speakers. Journal of Phonetics 29, 407–429.
Rodgers, Jonathan 1999. Three influences on glottalization in read and spontaneous German speech. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel 25. 173–280.
Siptár, P. 2002. Hiátus [Hiatus]. In Hunyadi, L. (ed.): Kísérleti fonetika – laboratóriumi fonológia 2002 [Experimental phonetics and laboratory phonology 2002]. Debrecen: Debreceni Egyetem Kossuth Egyetemi Kiadója. 85–97.
Siptár, P. 2008. Hiatus resolution in Hungarian: an optimality theoretic account. In Piñón, Ch. and Szentgyörgyi, Sz. (eds.): Approaches to Hungarian 10: Papers from the Veszprém Conference. Budapest: Akadémiai Kiadó. 187–208.
Siptár, P. 2012. The fate of vowel clusters in Hungarian. In Cyran, E., Szymanek, B. and Kardela, H. (eds.): Sound, structure and sense. Studies in memory of Edmund Gussmann. Lublin: Wydawnictwo KUL. 673–693.
Siptár, P. and Törkenczy, M. 2000. The phonology of Hungarian. Oxford: Oxford University Press.
Slifka, J. 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. Journal of Voice 20/2, 171–186.
Szende, T. 1973. Spontán beszédanyag gyakorisági mutatói [Frequency indices of

spontaneous Hungarian speech]. Nyelvtudományi Értekezések 81. Budapest: Akadémiai Kiadó.
Yuasa, I. P. 2010. Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women. American Speech 85/3, 315–337.

DERIVATION AND ACOUSTIC EFFECTS OF AN AREA FUNCTION FOR THE LARYNGEAL SUBGLOTTIS

Steven M. Lulich
Department of Speech and Hearing Sciences, Indiana University
e-mail: slulich@indiana.edu

Abstract

In speech and voice production, the larynx is a critically important organ. The larynx houses the vocal folds which vibrate to produce the voice source, and it also acts as a coupler between the acoustically resonant systems of the subglottal airways and the vocal tract. In this paper, one particular aspect of the larynx is investigated: the geometry and acoustic effects of the subglottis. The subglottis, defined as the portion of the airway between the roughly cylindrical trachea and the vibrating portion of the vocal folds, is found to have an area function which roughly follows that of the M5 model with a 40° divergence angle. From this result, which is based on the morphological study by Šidlof et al (2008), a model of the subglottis + subglottal airways acoustic input impedance is developed, and it is by found that the subglottis essentially acts as an acoustic mass, the effects of which are to lower the subglottal resonance frequencies a small amount, while the non-circular cross-section serves to increase the first resonance frequency as well as the compliance of the subglottal input impedance up to approximately 500 Hz.

1 Introduction

The geometry as well as the aerodynamic and acoustic effects of the laryngeal subglottis during phonation are poorly understood. On one hand, it is generally ignored in models of subglottal acoustics (cf. Fredberg and Hoenig, 1978; Habib et al., 1994 and Hudde and Slatky, 1989; Harper et al, 2001; Lulich, 2006, and Ho et al, 2011, all of which define subglottal input impedance from the top of a cylindrical trachea), and on the other hand, its geometry is arbitrarily simplified in many models of vocal fold vibration and voice production (cf. Scherer et al., 2001; Zañartu et al., 2007; Berry et al., 1994; Zhang et al., 2006). The author is not aware of any previous discussion of the acoustic effects of the precise subglottis geometry, except for the final paragraph of Titze (2008): “Perhaps the subglottal entry configuration can also be changed [to facilitate vocal fold vibration by increasing the acoustic] compliance.” On the other hand, if the subglottis can be approximated as a horn of circular cross-section, previous work has shown that such a horn can be modeled as a uniform tube with additional inertive shunt impedances at both ends (Fant, 1960; Benade, 1988), which might be expected to raise the resonant frequencies slightly when compared with a cylindrical tube of identical length.

Šidlof et al (2008) published results of a study in which they obtained three dimensional casts and images of the subglottis of pre-phonatory excised larynges. One particular larynx (larynx No. 8), with a pre-phonatory geometry corresponding to a fundamental frequency of 304 Hz, was chosen for more detailed study, and the results were given in numerical tables, which form the basis of the present study. Šidlof et al (2008) developed a non-destructive casting technique for determining the three dimensional geometry of excised human larynges. After the casts were digitized by a three dimensional scanner, coronal sections were extracted from two larynges, and a set of cubic polynomials was used to characterize each section. The information necessary to reproduce these cubic splines for larynx No. 8 was presented in tables for eight sections separated by steps of 1 mm. The goals of the present study are to derive an area function of the subglottis on the basis of the spline data presented by Šidlof et al (2008), and to determine the acoustic effects of such an area function on the subglottal input impedance.

In Section 2 of this paper, the methods employed to derive the area function for the subglottis are presented in detail. In Section 3, the acoustic effects of the subglottis are investigated. Section 4 presents a summary and conclusion.

2 Derivation of the subglottis area function

The goal of this section is to map three-dimensional measurements of the subglottis of an excised larynx to a simple area function for the same subglottis. Šidlof et al (2008) published tables of spline coefficients for the two vocal folds of larynx No. 8 at each of eight equally-spaced locations along the anterior-posterior axis. This section presents the methods employed to derive an area function from these published spline data. Although the morphologic study of Šidlof et al (2008) presents the most complete set of three-dimensional data on the subglottis geometry that the author is aware of, the derivation of an area function from these data is still subject to numerous sources of error, which will be discussed below. It is important to note, therefore, that the proposed area function is an estimate only.

A coordinate system was first established within which the analysis of the spline data was performed. According to this coordinate system, the superior-inferior direction defined the x-axis, the medial-lateral direction defined the y-axis, and the anterior-posterior direction defined the z-axis, as illustrated in Figure 1. Because the spline coefficients published by Šidlof et al (2008) were given with a spatial resolution of 10^{-6} mm, the same resolution was used here. This ensured that the junctions between different spline sections had the same spatial resolution as the points within the individual splines. Since this paper is concerned only with the subglottis, only the corresponding α, β, and γ spline sections published by Šidlof et al (2008) were used.

Axial cross-sections of the spline data were then examined. In order to identify the glottis, the distances between corresponding points on the left and right vocal folds were summed, and the x-coordinate for which this sum was smallest was defined to be that of the glottis. It became clear that the natural axis of the glottis

was not exactly in line with the defined anterior-posterior axis. Therefore, a new "glottal axis" was defined by fitting a straight line (of the form z = my + b) to the points midway between the left and right vocal folds.

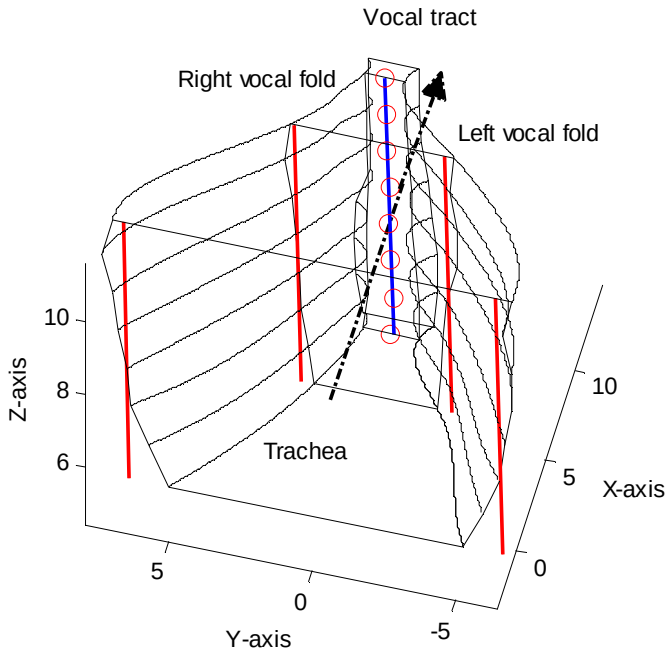


Figure 1. Definition of the coordinate system used in this paper and illustration of the process for deriving the area function for larynx No. 8 from Šidlof et al (2008). The dotted line ending in an arrow indicates the direction of airflow parallel to the x-axis. At all levels (along the inferior-superior x-axis), the cross-sections are approximately rectangular. The red open circles indicate the points midway between corresponding points on the left and right vocal folds. The blue line through the red circles is the best fit line characterizing the glottal axis. The red lines are parallel to the glottal axis but approximate the edges of the vocal folds at different locations along the x-axis.

Since the axial cross-sections of the spline data appear to be roughly rectangular at all levels along the x-axis, it was determined that the edge of each vocal fold could be approximated by a straight line parallel to the glottal axis. Each such line (of the form $z = my + b_l$ or $z = my + b_r$, where b_l and b_r refer to the z-intercept of the lines approximating the left and right vocal folds, respectively) was determined such that the summed error was zero, i.e. $\Sigma(y - [z - b_i]/m) = 0$, where the subscript i refers to either the left or right vocal fold. This ensures that the area between each vocal fold and the glottal axis is identical whether the original values of y or the approximated values $(z-b_i)/m$ are used. This result, however, is dependent on the spatial resolution of the data along the z-axis, which in this case is relatively poor

(only eight data points), and the true axial cross-section geometry at each level along the x-axis. For the levels near the glottis, the rectangular cross-section appears to approximate the data well. For the more inferior levels, on the other hand, the rectangular cross-section resembles a square though it may seem to be more accurate to model this cross-section as a circle since the tracheal cross-section is roughly circular. Therefore the area function to be derived from the rectangular cross-sections is presumably more accurate near the glottis and less accurate inferiorly near the trachea.

The x-coordinates of the inferior-most and superior-most points of each vocal fold spline were not constant across the eight locations, as shown in Figure 2. Among the inferior-most points, the largest x-coordinate was identified and the remaining splines were truncated at this value. Similarly for the superior-most points, the smallest x-coordinate was identified and the remaining splines were accordingly truncated. This ensured that every spline extended over the same range of x-values.

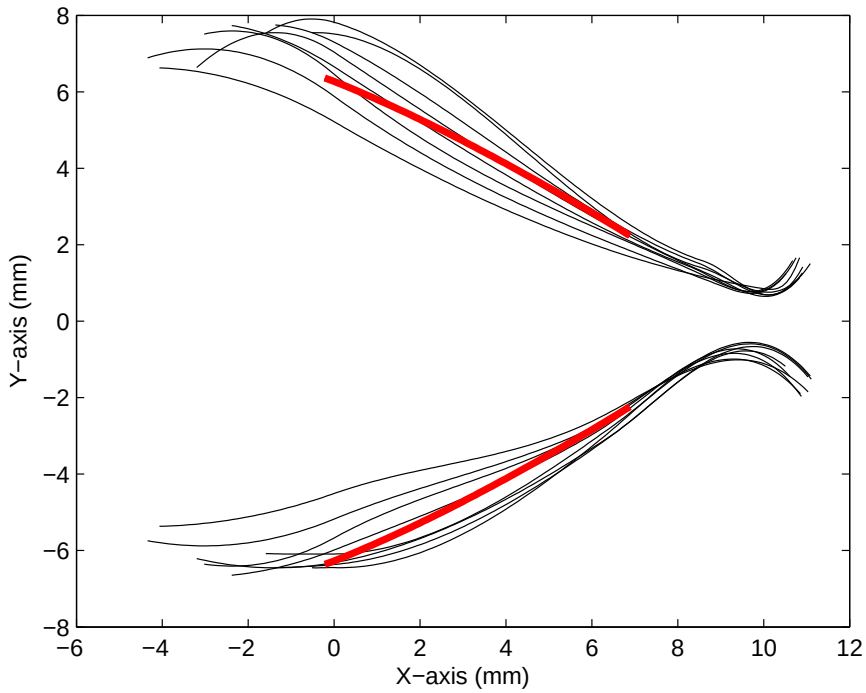


Figure 2. Vocal fold contours for each of eight locations along the anterior-posterior axis: the x-coordinates of the contour endpoints are not identical from one contour to the next. This is especially clear for the inferior endpoints. The red lines indicate the equivalent contours for a rectangular cross-section derived in this paper and truncated at both the inferior and superior ends as described in the text.

Once the parallel line segments approximating the vocal fold edges were determined, it was necessary to define the distance between them. In the present

study, this was defined as the length, d, of the line segment intersecting the two edges at right angles. The vocal folds were assumed to have an anterior-posterior length of 16 mm, so that the cross-sectional area at each location along the x-axis was defined to be $A = 16d$ mm^2. The area function was thus determined as a function of x. The assumption that the vocal folds have a length of 16 mm is consistent with a tracheal radius of 8 mm (Horsfield et al, 1971) and a corresponding cross-sectional area of the trachea equal to 64π mm^2.

The area function of the subglottis was therefore truncated at the point where it had the same area as the trachea. Furthermore, it was assumed that the portion of the vocal folds less than 3 mm inferior to the glottis oscillates during phonation with amplitude sufficiently large to be separated from the remainder of the subglottis (cf. Fig. 5 of Berry et al, 2001). Therefore, the area function was truncated at the point 3 mm inferior to the glottis. The resulting truncated area function yields the red lines in Figure 2, where the distance between the upper and lower red lines at a given x-coordinate is equal to the corresponding value of d. The length of the resulting subglottis area function is 7.48 mm.

Models of vocal tract and subglottal acoustics typically rely on the assumption that the cross-sectional shape of the airways is circular. Therefore, it was deemed useful to define a circular geometry at each level along the x-axis with an area equivalent to that determined above using the rectangular geometry. The radius as a function of x was therefore determined as $r = (A/\pi)^{1/2}$. Acoustic inertance and compliance are functions of cross-sectional area and therefore equivalent whether a circular or rectangular cross-section is assumed. Acoustic resistance and conductance, however, as well as vibrations and dissipation in the walls, are dependent on the airway cross-section perimeter. In the present case, the ratio of rectangular perimeter to circular circumerance (where the rectangle and circle have equal areas, and the rectangle has one side 16mm long) ranges from 10% (at the tracheal junction) to 35% (at the junction with the vibrating vocal folds). Any acoustic losses modelled using the circular geometry will therefore include some amount of error, a point which is explored in the next section.

As shown in Figure 3, the radius function derived here is similar to that of the 40°-M5 model (Scherer et al., 2001). In contrast, the model of Zhang et al., 2006 (30°-M5 model) shows a more rapidly increasing radius and thus a shorter subglottis. The quadratic model (Zañartu et al., 2007) results in a subglottis length similar to the 40°-M5 model and the model derived here, but with a larger discontinuity at the junction with the trachea and a smaller expansion of the radius function near the glottis.

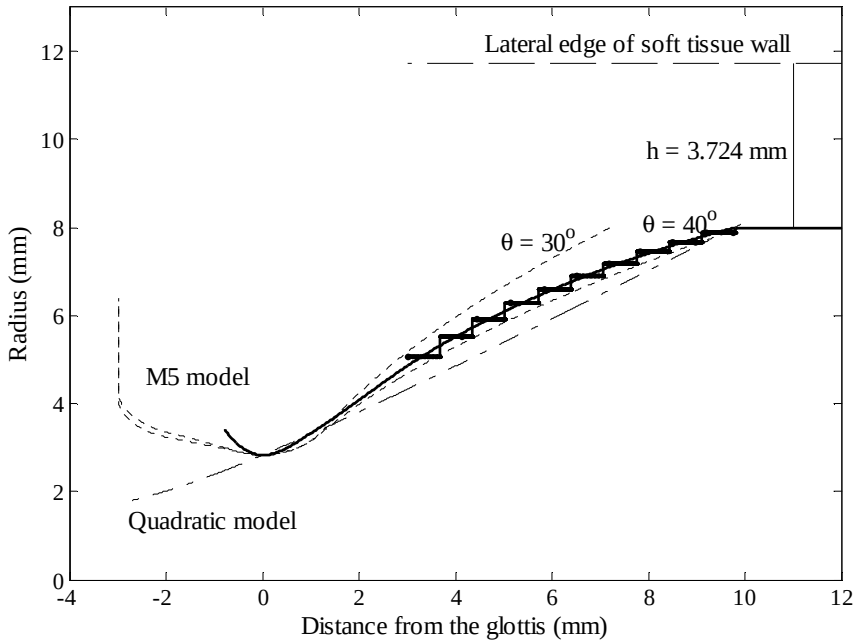


Figure 3. Illustration of the segmentation of the subglottis radius function: The radius of each uniform tube segment is chosen such that the cross-sectional area of the tube is equal to the average cross-sectional area of the subglottis area function within the same x-coordinate bounds. The outer boundary of the soft tissue wall is also shown as a dashed line. For the trachea, the wall thickness is 3.724 mm. The extension of the radius function in the direction of both the oscillating part of the vocal folds and the trachea is also shown. For comparison, two versions of the M5 model (Scherer et al., 2001) are shown (with 30° and 40° divergence angles), as well as a quadratic model (Zañartu et al., 2007; Berry et al., 1994). The 30°-M5 model has the same subglottis area function as the model used by Zhang et al., 2006).

3 Acoustic effects of the subglottis

The subglottal airways are modeled using the lung geometry of Horsfield et al (1971), and the input impedance is determined by modeling each airway as a continuous transmission line as described by Hudde and Slatky (1989). Mechanical properties of the airway wall tissues are assumed to be those presented by Harper et al (2001), with the exception that the soft tissue Young's modulus is assumed to be one order of magnitude larger than previously reported (i.e. $E_{ws} = 3.92 \cdot 10^6$ dyne/cm^2 rather than $3.92 \cdot 10^5$ dyne/cm^2). This modification was chosen since recent work has indicated that the resonance frequency of the soft tissues is probably significantly closer to the first subglottal resonance than previously thought (Lulich et al, 2011), and the most likely cause of this increased resonance frequency is an increase in the stiffness of the soft tissues when stretched in response to the phonatory subglottal pressure (Gunst and Stropp, 1988).

The subglottis itself does not have a constant cross-sectional area, so the best means by which its acoustic properties should be modeled remains an open question. One possibility is to approximate the subglottis area function as that of a conical horn (Benade, 1988). However, Benade's study applies only to lossless horns with rigid walls, and the present author is unaware of any subsequent study deriving the acoustic properties of horns with acoustic losses and yielding walls. An alternative approach is to divide the subglottis into a number of segments, each of which can be approximated by a short uniform tube (following Fant, 1960) (see Figure 3). This is the approach adopted for the present study.

For each segment, the mechanical properties of the soft tissue walls were assumed to be identical to those of the trachea, with the exception that the wall thickness was assumed to increase from the trachea to the glottis, as shown in Figure 3. Since the distance between the medial vocal fold edge and the laryngeal cartilages increases quickly from the trachea to the glottis, the subglottis wall impedance was considered to be purely due to soft tissues. The walls of the trachea and other bronchial airways were modeled using both soft tissues and cartilage in parallel, as described by Habib et al., (1994).

Since the number of uniform tube segments fit to the subglottis area function will affect the acoustic properties of the model, this number was varied to include 1-segment, 2-segment, and 10-segment models of the subglottis. It was found that the acoustic effects of the 2-segment and 10-segment models were virtually indistinguishable. The resulting input impedances of 1-segment, 2-segment, and 10-segment models of the subglottis alone, and that of the subglottis in series with the subglottal airways, is shown in Figure 4. In all cases, each segment of the subglottis was assigned a cross-sectional area equal to the mean value of the area function over the corresponding range. For example, the 1-segment model was assigned a cross-sectional area equal to the mean value of the entire 7.48 mm long area function.

To the extent that the derived 10-segment subglottis area function is approximately accurate, and that the wall mechanical properties and acoustic losses are adequately modeled, the effect of the subglottis on the input impedance of the subglottal airways is essentially to lower the frequency of each resonance by a factor of approximately 0.97 (23, 48, and 54 Hz for the first three resonances, respectively), and to increase the amplitude of each resonance.

The 2-segment and 10-segment subglottis models have less effect on the resonances than the 1-segment model. As the number of segments decreases, the more uniform the subglottis area function becomes, and the better it can be approximated as a lumped acoustic mass in series with the trachea. This accounts in the 1-segment model for both the decreased resonance frequencies and the high-pass characteristic leading to higher amplitudes at increasingly higher frequencies. For the 2-segment and 10-segment models, the resulting input impedance is intermediate between that of the 1-segment model and that of the subglottal airways without any subglottis. This bears out the prediction that the horn-shaped subglottis, represented as a uniform transmission line with primarily inertive shunt impedances (cf. Benade,

1988; Fant, 1960), should give rise to higher resonance frequencies than a uniform tube of identical length (i.e. the 1-segment model).

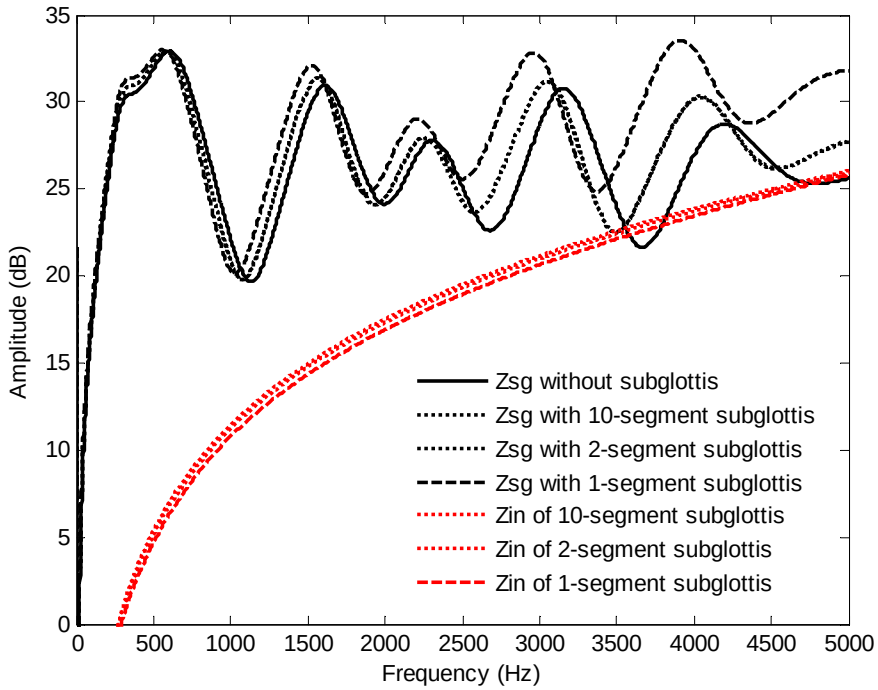


Figure 4. Input impedances: The black solid line shows the input impedance of the subglottal airways using the Horsfield lung geometry with no subglottis attached. Three additional input impedances are shown for the Horsfield geometry with 1-segment (black dashed), 2-segment (black dotted), and 10-segment subglottis models (black dotted). The input impedance of the subglottis alone is also shown for 1-segment, 2-segment, and 10-semgent subglottis models (red lines). The difference between the 2-segment and 10-segment models is small for both the subglottis alone and for the subglottis plus Horsfield geometry.

As noted above, acoustic losses and wall vibrations are dependent on airway cross-sectional perimeter rather than on airway cross-sectional area. For a given cross-sectional area, either a circular or a rectangular cross-section can be defined. For the circle, the perimeter is $2\pi r$, while for the rectangle it is $32 + 2d$, where 32 is twice the anterior-posterior length of the glottis and d is the width as defined above in Section 2. For each segment, the perimeter 32+2d was calculated and substituted for the circumference used to calculate the wall lumped elements as well as the acoustic resistance and conductance. The wall elements and the acoustic resistance also depend on r^2 in the circular case. The corresponding value in the rectangular case was replaced by $[(32+2d)/(2\pi)]^2$. Figure 5 shows the results.

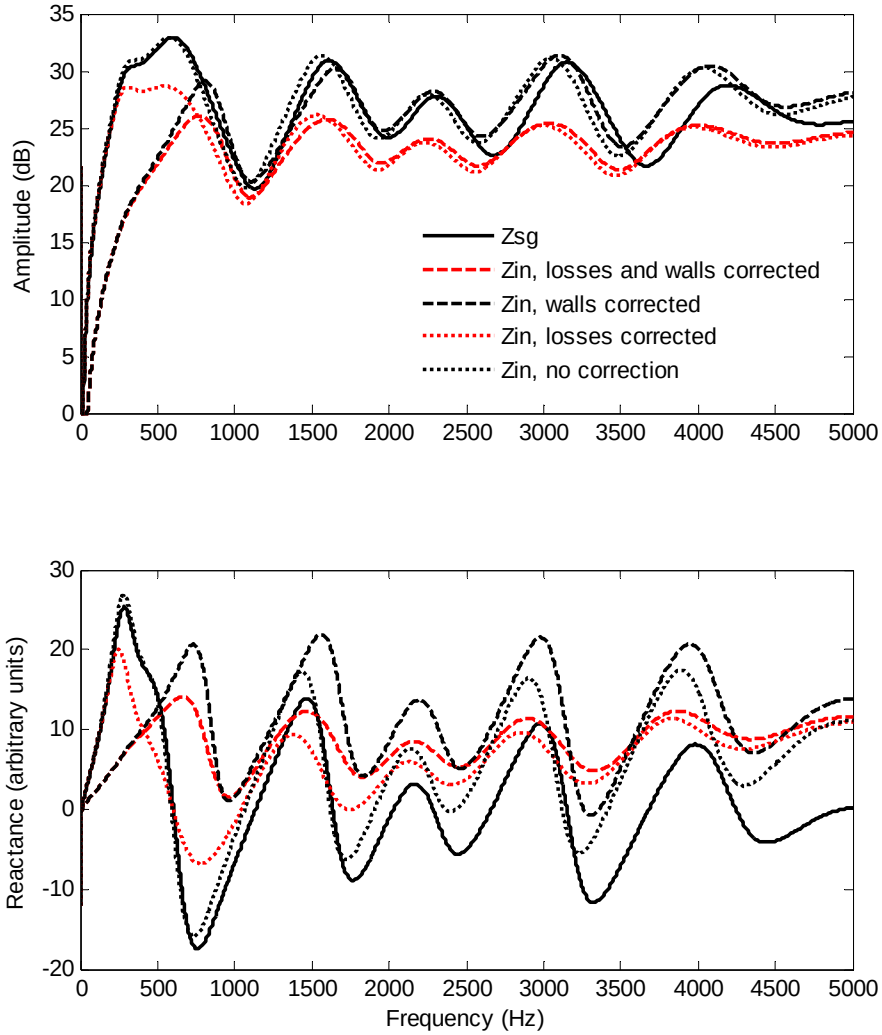


Figure 5. Top panel: Input impedances: The solid black line shows the input impedance of the subglottal airways with no subglottis attached. The remaining lines show the input impedance with the 10-segment subglottis model attached in series and with perimeter corrections applied to varying degree (see the legend). Bottom panel: Reactance (imaginary part) of the impedances shown in the top panel. Compliance and reactance are reciprocal, so that smaller reactance values correspond to larger compliance values.

When these approximate corrections are applied to both the acoustic losses and the wall mechanics (cf. Figure 5, red dashed line), the input impedance of the 10-segment subglottis model in series with the Horsfield geometry shows markedly reduced amplitudes of the resonances as well as a substantial increase in the frequency of the first resonance and a less prominent tissue resonance below 500 Hz. Moreover, the reactance is substantially smaller (more compliant) up to about

500 Hz, which covers the typical range of the fundamental frequency during speaking. When the perimeter correction is applied only to the acoustic losses, the result is much the same except that in the low frequency range the compliance is more similar to the uncorrected case, and the tissue resonance is more prominent. When the correction is applied only to the wall mechanics, the compliance remains high at low frequencies, while at high frequencies the reactance is more massive and the impedance amplitude is similar to that of the uncorrected input impedance. All of this suggests that Titze's conjecture about the subglottis "entry configuration" may be correct: a more rectangular configuration facilitates phonation while a more circular configuration hinders phonation. Moreover, it is apparently the wall tissue impedance that principally contributes to the increased compliance in the range of typical fundamental frequencies. It will therefore be of some interest in the future to determine more precisely what the mechanical properties are for the subglottis walls (e.g. Young's Modulus, viscosity, and thickness of the walls).

4 Summary and Conclusion

In this paper, an area function for the laryngeal subglottis was derived from morphological data previously published by Šidlof et al (2008). The subglottis is roughly rectangular near the glottis and becomes more square or circular where it joins with the trachea. Since a rectangular shape was assumed for the entire subglottis in this study, it is likely that the area function is somewhat less accurate as distance from the glottis increases. However, for practical purposes in which only the frequency range of the first three subglottal resonances is of interest, the acoustic effects of the subglottis appear to be relatively independent of the exact area function. For instance, if the area function is modeled as a set of concatenated uniform tubes, the acoustic effects of using more or fewer tubes are negligible.

The calculation of subglottis acoustics in this study initially assumed that the cross-sectional shape of the subglottis is circular throughout, but with the correct equivalent cross-sectional area. The assumption of a circular cross-section does, however, affect the cross-section perimeter, which influences the determination of acoustic losses and of the vibration of the wall tissues. Applying an approximate correction to account for the non-circular perimeter revealed that the input impedance was strongly dependent on the cross-section shape, especially at frequencies below 500 Hz, where the compliance increased for the rectangular cross-section. This increased compliance may have important effects on voice production.

Of the various previous models compared with the area function derived here, the 40°-M5 model (Scherer et al, 2001) appears to offer the best match. However, it is not known how well the derived area function will generalize to other excised larynges, phonation frequencies, or, most importantly, to living humans. The precise results presented here must therefore remain rough estimates. A method for deriving an area function from three-dimensional measurements of the left and right vocal fold shapes in pre-phonatory position was presented, which in this case led to a close

approximation of the 40°-M5 model. Regardless of the precise area function, the main acoustic effects appear to be to decrease the resonance frequencies of the subglottal input impedance, to raise the first subglottal resonance frequency, and to increase the compliance of the subglottal input impedance for frequencies below approximately 500 Hz. These main effects are primarily due to the overall length of the subglottis and its non-circular cross-section shape – two factors which do not significantly depend on the precise area function. In the future, it will therefore be most important to determine these two factors.

5 Acknowledgements

This work was supported in part by NSF Grant No. 0905250.

References

Benade, A. H. 1988. Equivalent circuits for conical waveguides. *Journal of the Acoustical Society of America* 83, 1764-1769.

Berry, D. A., Montequin, D. W. and Tayama, N. 2001. High-speed digital imaging of the medial surface of the vocal folds. *Journal of the Acoustical Society of America* 110, 2539-2547.

Fant, G. 1960. *Acoustic Theory of Speech Production, with Calculations based on X-Ray Studies of Russian Articulations.* The Hague, Mouton.

Fredberg, J. J and Hoenig, A. 1978. Mechanical response of the lungs at high frequencies *Journal of Biomechanical Engineering* 100, 57-66.

Gunst, S. J. and Stropp, J. Q. 1988. Pressure-volume and length-stress relationships in canine bronchi in vitro. *Journal of Applied Physiology* 64, 2522-2531.

Habib, R. H., Chalker, R. B., Suki, B. and Jackson, A. C. 1994. Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans. *Journal of Applied Physiology* 77, 441-451.

Harper, V. P., Kraman, S. S., Pasterkamp, H., and Wodicka, G. R. 2001. An acoustic model of the respiratory tract. *IEEE Transactions on Biomedical Engineering* 48, 543-550.

Ho, J. C., Zañartu, M., and Wodicka, G. R. 2011. An anatomically-based, time-domain acoustic model of the subglottal system for speech production. *Journal of the Acoustical Society of America* 129, 1531-1547.

Horsfield, K, Dart, G., Olson, D. E., Filly, G. F. and Cumming, G. 1971. Models of the human bronchial tree. *Journal of Applied Physiology* 31, 207-217.

Hudde, H. and Slatky, H. 1989. The acoustical input impedance of excised human lungs – measurements and model matching. *Journal of the Acoustical Society of America* 86, 475-492.

Lulich, S. M. 2006 The role of lower airway resonances in defining vowel feature contrasts. Ph.D. thesis, MIT.

Lulich, S. M., Alwan, A., Arsikere, H., Morton, J. R. and Sommers, M. S. 2011. Resonances and wave propagation velocity in the subglottal airways. *Journal of the Acoustical Society of America* 130, 2108-2115.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. 2001. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *Journal of the Acoustical Society of America* 109, 1616-1630.

Šidlof, P., Švec, J. G., Horáček, J., Veselý, J., Klepáček, I. and Havlík, R. 2008. Geometry of human vocal folds and glottal channel for mathematical and biomechanical modeling of voice production. *Journal of Biomechanics* 41, 985-995.

Titze, I. R. 2008. Nonlinear source-filter coupling in phonation: Theory. *Journal of the Acoustical Society of America* 123, 2733-2749.

Zañartu, M., Mongeau, L., and Wodicka, G. R. 2007. Influence of acoustic loading on an effective single mass model of the vocal folds. Journal of the Acoustical Society of America 121, 1119-1129.

Zhang, Z., Neubauer, J., and Berry, D. A. 2006. The influence of subglottal acoustics on laboratory models of phonation. *Journal of the Acoustical Society of America* 120, 1558-1569.

ACOUSTIC ANALYSIS OF FORMANT SHIFTS IN NASALIZED VOWELS

Takayuki Arai
Dept. Information and Communication Sciences, Sophia University
e-mail: arai@sophia.ac.jp

Abstract

We studied the formant frequency shifts of nasalized vowels. Based on acoustic theory, the first formant (F1) increases for high vowels, while F1 decreases for low vowels. In the present study, we measured formant frequencies for the following: 1) nasalized vowels produced by physical models of a vocal tract, and 2) nasalized vowels uttered in a nasal context. As predicted by acoustic theory and perceptual findings, acoustic analyses revealed bidirectional formant shifts in F1 frequency: increasing F1 for high vowels and decreasing F1 for low vowels.

1 Introduction

The nasal tract may couple to the main vocal tract for several reasons, including either an anatomical or functional problem (Stevens et al., 1986). Cleft palate patients, for example, often have velopharyngeal insufficiency that causes hypernasality, and one of the most common problems for deaf speakers is inadvertent nasalization, a speech disorder where the velopharyngeal port opens excessively during vowel production (Stevens et al., 1976; Chen, 1995). The ability to control coupling of the nasal cavities to the vocal tract is crucial for the production of normal speech (Bell-Berti, 1980). Inability to control coupling results in severely distorted speech.

Certain languages, such as French, Portuguese and Hindi, phonemically distinguish nasal and nonnasal vowels, whereas other languages, including English, do not. Even in languages where nasalization is not phonemic, nasal coupling occurs during the production of oral vowels adjacent to nasal consonants due to articulatory overlap of the velum and the tongue or lips. If the vowel is preceded by an obstruent and followed by a nasal consonant, the velopharyngeal port is closed at the time of release of the obstruent but opens during the vowel in preparation for the formation of the oral closure for the nasal consonant (Stevens, 1998). These overlapping gestures result in the preceding vowel being nasalized before a nasal consonant.

The degree of acoustic coupling between the vocal and nasal tracts is controlled by the velum as well as by the posterior and lateral pharyngeal walls (Skolnick et al., 1973). The opening to the nasal tract allows airflow through the nose and mouth, and acoustic coupling causes the vowel to be nasalized. Therefore, a simple model for a nasalized vowel could be a main vocal tract with a side branch, where the

degree of opening of the velopharyngeal port controls the degree of nasalization. According to acoustic theory (Fant, 1960; Fujimura, 1960, 1961; Fujimura and Lindqvist, 1971; House and Stevens, 1956) the basic difference between the transfer function for the vocal tract with a side branch and that for a nonnasal vowel is that additional poles and zeros are introduced to the vocal-tract transfer function as a consequence of acoustic coupling to the nasal tract. The additional poles and zeros due to nasal coupling cause modifications in the spectrum, such as reduction in amplitude of the first formant (F1), broadening the bandwidth of F1, shifting F1 upwards in frequency, and a relative strengthening of the spectrum near 250 Hz (House and Stevens, 1956; Hattori et al., 1958; Fant, 1960; Fujimura, 1960; Fujimura and Lindqvist, 1971; Hawkins and Stevens, 1985; Maeda, 1993). The higher frequencies may also be modified by nasal coupling. The main effect of nasalization, however, is the perturbation of the low-frequency spectrum by replacement of the first formant with a shifted F1 (F1'), a nasal formant (Fn), and a nasal zero (Fz) (Fant, 1960; Fujimura and Lindqvist, 1971; Stevens et al., 1986). As the cross-sectional area of the velopharyngeal opening is gradually increased, the spacing between the pole and zero introduced in the vicinity of the first formant increases, F1 frequency shifts, and F1 bandwidth increases.

Calculations of the acoustic consequences of nasal coupling predict distinctively different modifications depending on vowel identity (Fujimura, 1960; Fujimura and Lindqvist, 1971). The theory predicts that F1 shifts upwards in frequency for high vowels, and a nasal formant appears in the spectral valley between F1 and F2. For low vowels, F1 also shifts upward in frequency, but at the same time, F1 comes close to zero. This is because the first pole-zero pair is lower in frequency than F1 in the corresponding oral vowel, and as a result, F1 is weakened and seemingly split into two peaks (Stevens et al., 1986). At low degrees of coupling, the nasal pole (Fn) is almost canceled by the nasal zero (Fz), and in this case, Fn is not prominent. At higher degrees of coupling, however, Fn increases in prominence (Fujimura and Lindqvist, 1971), and Fn could be identified as a formant (Maeda, 1993).

Thus, nasal coupling can shift F1 frequency and this may affect perceived vowel height. Due to the upward shift in F1 frequency, nasalization might be expected to lower perceived vowel height (Ohala, 1986). In addition, due to the prominence of Fn, nasalization might be expected to raise perceived vowel height for low vowels. This appears to be a plausible explanation for the bidirectional shifts in perceived nasal vowel height (Krakow et al., 1988) observed in perceptual experiments (Wright, 1975, 1986).

This F1 shift might lead to listener misperceptions of vowel height. Krakow et al. (1988) examined the hypothesis that perceived nasal vowel height is not entirely determined by the spectral shape of the nasal vowel, but rather that the context in which the nasal vowel occurs can affect the way in which the nasalization of that vowel is perceived. They tested this hypothesis by comparing listeners' perception of nasal vowels in the presence and absence of an adjacent nasal consonant, finding

that nasal coupling does not necessarily lead to listener misperceptions of vowel quality when the vowel's nasality is coarticulatory in nature.

In English and many other languages, vowels should be perceived as the same phoneme regardless of nasalization. In other words, a speaker and/or a listener might tend to compensate for any such formant shifts. Arai (2005) investigated whether compensation exists in vowel production. He measured the positions of the articulators, especially tongue height, and compared them in oral and nasal contexts using the electromagnetic midsagittal articulometer (EMMA) system (Perkell et al., 1992). The measurement of the positions of the articulators showed almost no compensation except for the lowest vowel /ɑ/.

The goal of this present study is to confirm through acoustic analyses the manner in which formant frequencies shift due to nasalization, particularly the bidirectional shifts in F1 frequency. Even though these bidirectional movements have been demonstrated by both acoustic theory and perceptual findings, clear acoustic evidence is scarce in the literature. Therefore, we aim to confirm that the F1 frequency tends to shift in a more central direction when it is nasalized, based on several measurements of formant frequencies for various vowels.

2 Theoretical considerations

The shift in formant frequencies due to nasal coupling during vowel production can be predicted by the acoustic theory of nasalized vowels. When the main vocal tract is in a vowel-like configuration, the velopharyngeal opening introduces poles and zeros into the transfer function of the vocal tract. These additional poles and zeros are approximated by applying an electric-circuit analog based on a simple model (Fujimura, 1960; Fujimura and Lindqvist, 1971; Stevens, 1998). In this model, a side branch is attached to the main vocal tract, and the susceptances looking in different directions from the velopharyngeal port are examined. Based on the measured transfer function of the nasal tract from above the closed velopharyngeal port to the nostril output, as measured using a sweep-tone source (Lindqvist and Sundberg, 1972), the susceptance looking into the nose is estimated to have zeros at about 500 and 2000 Hz (Chen, 1997; Stevens, 1998).

This model predicts how F1 is replaced by F1' where Fn and Fz depend on vowel height (Stevens, 1998). For all vowel types, as the area of the velopharyngeal port increases, an extra pole-zero pair (Fn and Fz) starts to appear near 500 Hz and shifts upwards, and the spacing of the pole-zero pair widens. An increase in coupling also corresponds to an upward shift in F1 frequency. For the vowel /i/, as the degree of nasal coupling continuously increases, F1' gradually shifts upwards from F1 of the non-nasal vowel and reaches a frequency lower than 500 Hz; Fn gradually shifts upwards from 500 Hz and reaches a frequency lower than 1 kHz; and Fz shifts upwards from 500 Hz to a frequency higher than 1 kHz. As a result, an extra pole occurs in the spectral valley between F1 and F2 in the spectrum of the nasalized vowel /i/. For the vowel /ɑ/, as the degree of nasal coupling continuously increases, the situation is a little more complicated (Stevens, 1998; Maeda, 1993). F1'

gradually shifts upwards from F1 of the non-nasal vowel, but at the same time, Fz rapidly shifts upwards from 500 Hz. Fn gradually shifts upward from 500 Hz and reaches a frequency lower than F1 of the non-nasal vowel. As a result, F1' becomes weakened by Fz, and Fn gradually becomes dominant. To restate, F1' first shifts upwards; then the F1 region has two peaks of F1' and Fn; and finally, Fn acts as F1, as Fn is dominant.

3 Acoustic analyses

As mentioned above, acoustic theory shows that F1 tends to shift in a more central direction when nasalized. To confirm the formant shifts, especially the bidirectional shifts in F1 that are due to nasalization, we conducted acoustic analyses on two sets of vowels: 1) mechanically produced nasalized vowels, and 2) naturally produced vowels in nasal and non-nasal contexts.

3.1. Mechanically produced nasalized vowels

To clearly observe the poles and zeros, we first made physical models of the human vocal tract for [i] and for [ɑ] with the nasal cavity (Arai, 2007), as shown in Fig. 1 (below). This model is designed to have less acoustic loss than the human vocal tract so that peaks with lower bandwidth can be obtained. Each of the models is made from five acrylic plates. The center (black) plate is 1 cm thick and has a schematic midsagittal cross-section for each vowel. On both sides of the center plate are two transparent 3-cm thick plates that have holes to achieve the proper area functions for the vocal-tract configuration of the vowels [i] and [ɑ] with nasal cavities. The outer-layered plates are 1 cm thick and are also transparent.

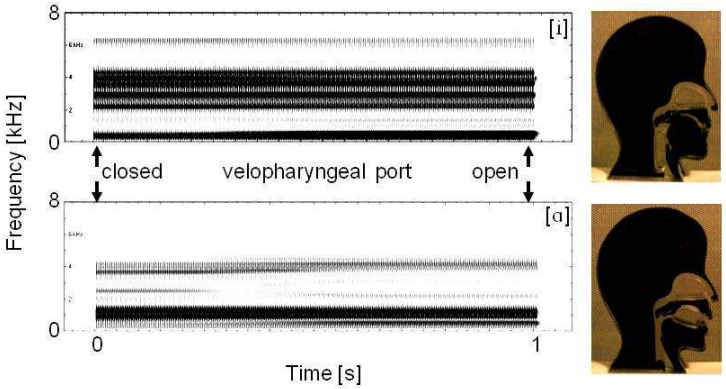


Figure 1. Physical models of the human vocal tract with the nasal cavity (right) and spectrograms of vowels produced from the physical models (left). The first panel: vowel [i]; and the second panel: vowel [ɑ].

The velum is made of rubber and may be rotated around a pivot located roughly at the boundary of the soft and hard palates. This movable velum acts as the velopharyngeal port and allows the simulation of different degrees of nasal coupling. The velopharyngeal opening is controlled by the rotating valve.

3.1.1 Method

We mechanically produced oral and nasalized vowels using the physical models of the human vocal tract. Default settings for the KLGLOTT88 voicing source model were used as a sound source (Klatt and Klatt, 1990). The radiation characteristics are already incorporated in this voicing source model. The fundamental frequency was set at 100 Hz. The velopharyngeal port was first closed, then opened mid-utterance. The final areas of the velopharyngeal port were approximately 30 mm^2 for [i] and 50 mm^2 for [ɑ].

The sound source was played from a laptop computer via the digital-to-analog (D/A) converter of a digital audio amplifier (MA-500U, Onkyo) that was connected to a laptop computer via a USB interface. The sampling frequency was 16 kHz. The amplifier then drove a driver unit (TU-750, TOA) used for a horn speaker. To avoid unwanted coupling between the neck and the area behind the neck of the driver unit and to achieve high impedance at the glottis end, we inserted a close-fitting hard rubber cylindrical filler inside the neck. We made a hole in the center of the rubber filling with an area of 30 mm^2. The neck part was attached to the glottal end of the vocal tract models.

The utterances were recorded using an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was placed approximately 20 cm in front of the model's lips in a partially sound-attenuated room, where the distances from the microphone to the mouth and to the nose were about equal. All signals were digitized with a sampling frequency of 16 kHz.

3.1.2 Results and discussion

The velum has been reported to be positioned higher in high vowels and lower in low vowels (Bell-Berti, 1980). To achieve the same degree of perceived nasality requires a small opening for high vowels but a greater velopharyngeal opening for low vowels (House and Stevens, 1956). Thus, we used different final areas of the velopharyngeal opening for the vowels [i] and [ɑ] at approximately 30 mm^2 and 50 mm^2, respectively.

Figure 1 shows the spectrograms for [i] and [ɑ]. For the vowel [i], the bandwidth of F1 became wider when the velopharyngeal port opened, and simultaneously, there was an upward shift of the F1 frequency. Furthermore, the extra pole Fn between F1' and F2 was observable around 1 kHz. For the vowel [ɑ], a pole-zero pair associated with the velopharyngeal opening was observed below the original F1 frequency. As the velopharyngeal opening widened, the frequencies of the pole and zero increased and, simultaneously, the distance between the pole and zero also increased. Eventually, the extra pole below the original F1 became dominant, as predicted by acoustic theory.

3.2 Naturally produced vowels in oral and nasal contexts

To investigate the formant shifts of natural vowels in real speech, we analyzed a speaker's vowels in oral and nasal contexts.

3.2.1 Method

The speech samples were monosyllabic nonsense words "bVC" uttered by a native, male speaker of American English. The vowel V was either /i/, /ɪ/, /ɛ/, /ʌ/, /æ/, or /ɑ/; and the consonant C was either /b/ or /m/. The target words were embedded in the carrier phrase "Say _____, again." All 12 combinations were repeated five times in random order (60 utterances in total). The utterances were recorded in a partially sound-attenuated room with an Electro-Voice model 054 omnidirectional dynamic microphone and a pre-amplifier (Shure Professional Microphone Mixer). The microphone was placed approximately 20 cm in front of the speaker's lips to establish equal distances from the microphone to the mouth and to the nose. All signals were digitized with a sampling frequency of 16 kHz. Formant tracking of the vowels was done by the LSPECTO program in XKL, which is a revision of the software package developed by Klatt (1984). The formant-tracking algorithm was based on 20th order linear predictive coding (LPC). For each frame, a 25-ms Hamming window was used, and each frame output was generated every 5 ms.

3.2.2 Results and discussion

Figure 2 shows the results of the formant tracking (formant frequencies versus time) for F1 of the six vowels. Each plot contains the results of the formant tracking of each /bVb/-/bVm/ pair. (We refer to the first one of the pair as "oral context", and the second as "nasal context.") This pairing allows us to isolate the effect of nasalization, because the only difference between the pair is whether or not there is velopharyngeal opening; the rest of the articulatory movements are identical (Krakow and Huffman, 1993). In each plot, the dots '.' and 'x' represent the formant frequency of either F1 or F2 at a particular time in /bVb/ and /bVm/, respectively. The thick and thin lines are the average curves obtained by smoothing with a 3rd-order polynomial approximation among five repetitions of /bVb/ and /bVm/.

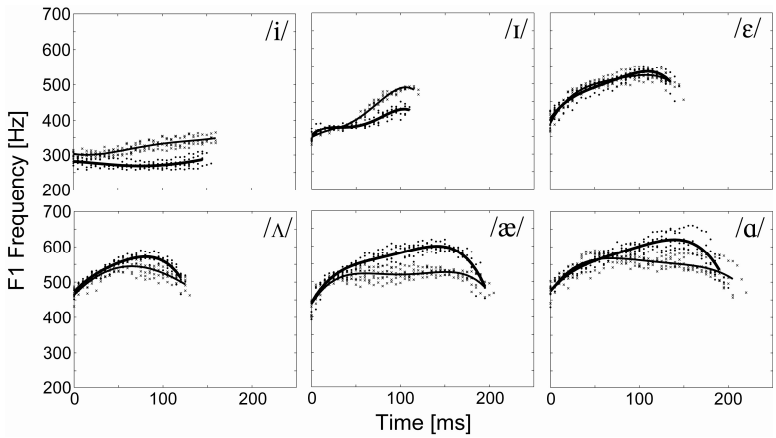


Figure 2. Physical models of the human vocal tract with the nasal cavity (right) and spectrograms of vowels produced from the physical models (left). The first panel: vowel [i]; and the second panel: vowel [ɑ].

From these plots in Fig. 2, we observed the bidirectional shifts in F1 due to nasalization, that is, the F1 frequency shifts downward for the low vowels /ɑ/ and /æ/ and upward for the high vowels /i/ and /ɪ/. However, there were no significant differences in F2 frequency among the six vowels.

To model the bidirectional shift in F1 frequency due to nasalization, we measured the maximum difference between the two average F1 curves in oral and nasal contexts (see Table 1). This maximum difference as a function of F1 frequency in an oral context can be modeled by a sigmoidal function (1).

Table 1. F1 frequencies of V in oral (/bVb/) and nasal (/bVm/) contexts at the point of maximum disparity.

V	F1 of (b)V(b) [Hz]	F1 of (b)V(m) [Hz]	Difference [Hz]
i	273.6	335.2	61.6
ɪ	420.7	483.4	62.8
ɛ	535.0	524.5	-10.5
ʌ	565.6	530.1	-35.5
æ	599.3	524.7	-74.7
ɑ	619.8	550.9	-68.8

$$\Delta F_1 = 67.8\left(\frac{2}{1+e^{0.0435(F_1-529.5)}}-1\right) \text{ [Hz]}. \quad (1)$$

From this model, we can see that the F1 frequency tends to shift bidirectionally toward the central region, which is around 530 Hz in this case. This coincides with the first natural frequency of the nasal cavity itself.

4 Discussion

Many studies have observed a peak associated with nasalization, especially between 250 and 450 Hz (Hattori et al., 1958; Fujimura and Lindqvist, 1971; Lindqvist-Gauffin and Sundberg, 1976; Takeuchi et al., 1977; Maeda, 1982a; Dang and Honda, 1995; Stevens, 1998). Most of these studies noted the possibility that side cavities corresponding to the nasal sinuses produced such a peak. Even when there is no coupling with the sinuses, however, a vowel will be nasalized with nasal coupling. In fact, we observed clear nasalization with the physical models with no sinuses in Section 3.1.

Ushijima and Sawashima (1972) found that the velopharyngeal port area decreases as velar elevation increases. Moreover, Moll (1962), among a number of investigators, concluded that velar elevation is greater for high vowels than for low vowels. In other words, velum height is greatest for obstruents and decreases according to the following order: high vowels, low vowels, and nasal consonants (Bell-Berti, 1980). Complete closure of the port is not always required for normal "non-nasal" speech production. To establish admittances into the nasal, oral, and

pharyngeal branches at the velopharyngeal port, the speaker need only make the port sufficiently small so as to prevent the nasal branch from affecting the overall vocal tract transfer function for sonorants (Bell-Berti, 1980).

Bell-Berti (1980) found that speakers having a minimum velopharyngeal port area (critical port area) of less than about 30 mm^2 produced speech that was nearly normal, while those having a larger minimum port area produced speech judged as nasalized. These data also agreed with the results of Isshiki et al. (1968), who induced velopharyngeal incompetence in their subjects by placing polyvinyl tubes in their velopharyngeal ports and found the critical port area to be about 20 mm^2.

In experiments with synthesized speech, House and Stevens (1956) reported that listeners failed to judge any of their vowel stimuli produced with a port area of 25 mm^2 as "more nasal" than those produced with the port completely closed. However, high vowels produced with a port area of 71 mm^2 (the next larger area in their series) were judged as "more nasal" than those produced with the smaller area. For the low vowel [ɑ] with a port area of 71 mm^2, the listeners' judgment stayed as low as with the smaller area. Thus, listeners judged that a greater velopharyngeal opening was needed to produce a given level of nasality for the low vowels than for the high vowels (Hawkins and Stevens, 1985). A similar observation was reported by Maeda (1993) with synthesized nasal vowels using his simulation method (Maeda, 1982b). This is consistent with observations of the articulators by Moll (1960), Delattre (1968), and Benguerel and Lafargue (1981), showing dependency of velopharyngeal port area on vowel height.

Using the physical models of the vocal tract in Section 3.1, the area of the velopharyngeal port that yielded the same degree of nasality differed for the vowels [i] and [ɑ], at approximately 30 mm^2 for [i] and 50 mm^2 for [ɑ]. Although these areas are slightly smaller than those reported in previous studies, the mechanically produced nasal vowels were sufficiently "nasalized."

As described earlier, once a vowel is nasalized, F1 is replaced by F1', Fn, and Fz, which can be observed as two poles and a zero in low frequencies of the spectrum. In physical models of a vocal tract, the bandwidths of the peaks tend to be narrower, which seems to be due to minor acoustic loss. As a result, we can specify the poles and zeros in the spectrum relatively easily.

In the case of nasalized vowels in real human speech, spectral peaks are less "peaky." This appears to be due to some degree of acoustic loss. It was therefore hard to specify the extra poles and zeros and we were not always able to observe F1', Fn, and Fz. Nevertheless, some might be evident. For example, we found Fn around the 1-kHz region in Fig. 1 for [i], and this is consistent with the findings of previous studies (House and Stevens, 1956; Maeda, 1982c). We also observed Fn in the frequency range below F1 for [ɑ] as shown in Fig. 1. We could only find the nasal zero Fz in some cases, such as the Fz between Fn and F1' for the mechanically produced [ɑ] in Fig. 1. However, the zero itself is expected to have less perceptual relevance than the additional spectral prominence, as pointed out by Stevens (1998).

For naturally produced vowels in the oral and nasal contexts in Section 3.2, we compared the minimal pairs /bVb/-/bVm/. It is a common observation that vowels adjacent to nasal consonants are nasalized because of overlapping gestures of the velum and the tongue or lips. Such overlapping gestures occur both when a vowel is followed by a nasal consonant and when a vowel is preceded by a nasal consonant. However, the overlap is somewhat longer in the former case (Stevens, 1998).

If the vowel V is preceded by an obstruent C in the CVN context where N is a nasal consonant, the velopharyngeal port is closed at the time of the release of the consonant C; the port then opens during the vowel V in preparation for the formation of the oral closure for the nasal consonant N (Stevens, 1998). This avoids build up of intra-oral pressure due to the oral closure for the nasal consonant. Because it takes time to lower the velum, the velopharyngeal port starts opening before complete closure of the oral cavity. Ohala (1971) has reported greater nasal coarticulation effects in vowels preceding nasals than in vowels following nasals, and states that velar lowering begins as soon as elevation is no longer required for obstruent articulation. Moll and Daniloff (1971) also reported that movement toward the opening of the velar port began during articulator movement toward the first vowel in a CVN sequence. For any vowel, F1' increases when the vowel is nasalized. As F1' increased, Fn was observed in frequencies lower than F1 for the low vowel [ɑ] and Fn became dominant as the velopharyngeal port area increased (particularly in Section 3.1). This observation from the acoustic measurements of the present study is consistent with the prediction of acoustic theory, and it supports the explanation that the F1 of a low vowel shifts downward due to the dominant Fn. Although this "F1 transfer" from F1' to Fn is understood to occur in real speech, it is a little more difficult to see such an F1 transfer for /ɑ/ from the results in Section 3.2 because of the initial transition from the preceding obstruent (/b/). However, the difference between the average F1 contours obtained from the vowels in nasal and oral contexts allows us to see an initial upward movement, which might be evidence of the F1-transfer phenomenon. For high vowels, because there is no such F1 transfer, F1' monotonically shifts upward in frequency. As a consequence, the bidirectional frequency shifts of F1 can be observed.

The nasalization of vowels shows the "quantal nature" of speech (Stevens, 1972, 1989). Let us consider that the velum lowers from the raised position (the complete closure of the velopharyngeal port). Although the lowering speed is constant, the perceived nasality does not increase constantly. Especially for a low vowel, nasality does not increase unless the velopharyngeal opening reaches a certain area, as described earlier. In fact, the velopharyngeal port does not close completely when we produce a low vowel with no perceived nasality. Interestingly, we might actually be able to hear less nasality with a slight opening of the velopharyngeal port (Maeda, 1993). Thus, for a low vowel, a plateau exists where a small perturbation in the velopharyngeal opening does not influence the perceived nasality. This type of nonlinear relation between articulation and acoustics or perception is a part of the quantal nature of speech (Stevens, 2003).

The quantal aspects of speech can also be observed in the formant frequency shifts of nasalized vowels. Particularly in the case of a low vowel, the complicated F1 movement due to the "transfer" phenomenon can be observed during the transition from complete closure to some degree of opening at the velopharyngeal port. Before the F1 transfer occurs, the perceived nasality is low and F1 frequency is more stable. After the F1 transfer, the perceived nasality is higher and the F1 frequency is somewhat stable. This nonlinear relationship between articulation and acoustics is also part of the quantal nature of speech, and one might predict that a nasal vowel tends to be produced in such a stable region, especially in languages that have a phonemic distinction between nasal and oral vowels, such as French (Maeda, 1993).

5 Conclusions

In this study, we investigated the formant frequency shifts of nasalized vowels. From acoustic measurements, bidirectional formant shifts in F1 frequency were observed (ranging from about –75 Hz for low vowels to about +65 Hz for high vowels in Section 3.2), as predicted by acoustic theory. From measurements using the EMMA system in Arai (2005), we found that speakers of American English tend not to compensate for such an F1 frequency shift by adjusting tongue height, except when producing the lowest vowel /ɑ/. From our perceptual experiment (Arai, 2004), we conclude that the compensation effect occurs even in an isolated vowel when it has both nasalization and formant transitions, so that listeners are able to predict that the vowel is in a nasal context. This supports the findings by Krakow et al. (1988) that nasal coupling does not necessarily lead to listener misperceptions of vowel quality.

6 Acknowledgements

This study was carried out while I was a Visiting Scientist with the Speech Communication Group in the Research Laboratory of Electronics, Massachusetts Institute of Technology (Cambridge, MA, USA) from 2000 through 2004. I would like to thank all of the people who helped me in various ways, especially Kenneth N. Stevens, Joseph S. Perkell, Stefanie Shattuck-Hufnagel, Sharon Manuel, Janet Slifka, Helen Hanson, Majid Zandipour, Mark Tiede, other members of the Speech Communication Group at MIT, Bernard Gold of MIT Lincoln Laboratory, John J. Ohala of University of California, Berkeley, Setsuko Imatomi of Mejiro University, and Kyoko Takeuchi of Kokugakuin University. This research was supported in part by Grants-in-Aid for Scientific Research (21500841, 24501063) from the Japan Society for the Promotion of Science. Portions of this work were presented at the 147th Meeting of the Acoustical Society of America, New York, N.Y., in May 2004.

References

Arai, T. 2004. Formant shift in nasalization of vowels. *Journal of the Acoustical Society of America* 115(5), Pt. 2, 2541.

Arai, T. 2005. Comparing tongue positions of vowels in oral and nasal contexts. *Proceedings*

of Interspeech, 1033-1036.
Arai, T. 2007. Education system in acoustics of speech production using physical models of the human vocal tract. *Acoustical Science and Technology* 28(3), 190-201.
Bell-Berti, F. 1980. Velopharyngeal function: A spatial-temporal model. In N. J. Lass (ed.): *Speech and Language: Advances in Basic Research and Practice 4.* New York: Academic. 291-316.
Benguerel, A.–P. and Lafargue, A. 1981. Perception of vowel nasalization in French. *Journal of Phonetics* 9, 309-321.
Chen, M. Y. 1995. Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *Journal of the Acoustical Society of America* 98, 2443-2453.
Chen, M. Y. 1997. Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America* 102, 2360-2370.
Dang, J. and Honda, K. 1995. An investigation of the acoustic characteristics of the paranasal cavities. *Proceedings of International Congress of Phonetic Sciences* 1. 342-345.
Delattre, P. 1968. Divergences entre la nasalité vocalique et consonantique en français. *Word 24*, 64-72.
Fant, G. 1960. *Acoustic Theory of Speech Production.* The Hague: Mouton.
Fujimura, O. 1960. Spectra of nasalized vowels. *Research Laboratory of Electronics Quarterly Progress Report* No. 58, MIT. 214-218 (July 15, 1960).
Fujimura, O. 1961. Analysis of nasalized vowels. *Research Laboratory of Electronics Quarterly Progress Report* No. 62, MIT. 191-192 (July 15, 1961).
Fujimura, O. and Lindqvist, J. 1971. Sweep-tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America* 49, 541-558.
Hattori, S., Yamamoto, K. and Fujimura, O. 1958. Nasalization of vowels in relation to nasals. *Journal of the Acoustical Society of America* 30, 267-274.
Hawkins, S. W. and Stevens, K. N. 1985. Acoustic and perceptual correlates of the non-nasal- nasal distinction for vowels. *Journal of the Acoustical Society of America* 77, 1560-1575.
House, A.S. and Stevens, K. N. 1956. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders* 21, 218-232.
Isshiki, N., Honjow, I. and Morimoto, M. 1968. Effects of velopharyngeal incompetence upon speech. *Cleft Palate Journal* 5, 297-310.
Klatt, D. H. 1984. The new MIT speech VAX computer facility. In *Speech Communication Group Working Papers* IV, Research Laboratory of Electronics, MIT, Cambridge. 73-82.
Klatt, D. H. and Klatt, L. C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820-857.
Krakow, R. A., Beddor, P. S., Goldstein, L. M. and Fowler, C. A. 1988. Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America* 83, 1146-1158.
Krakow, R.A. and Huffman, M.K. 1993. Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction. In Huffman, M. K. and Krakow, R. A. (eds.): *Phonetics and Phonology* 5. San Diego: Academic Press,. 3-59.
Lindqvist, J. and Sundberg, J. 1972. Acoustic properties of the nasal tract. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1, 13-17.
Lindqvist-Gauffin, J. and Sundberg, J. 1976. Acoustic properties of the nasal tract. *Phonetica* 33, 161-168.
Maeda, S. 1982a. The role of the sinus cavities in the production of vowels. *Proc. of International Conference on Acoustics, Speech, and Signal Processing.* Paris. 911-914.
Maeda, S. 1982b. A digital simulation method of the vocal-tract system. *Speech Communication* 1. 199-229.
Maeda, S. 1982c. Acoustic cues for vowel nasalization: A simulation study. *Journal of the*

Acoustical Society of America, 72(S1), S102.
Maeda, S. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In Huffman, M. K. and Krakow, R. A. (eds.): *Phonetics and Phonology: Nasals, Nasalization, and the Velum* 5. San Diego, CA: Academic Press. 147-167.
Moll, K. L. 1960. Cinefluorographic techniques in speech research. *Journal of Speech and Hearing Research* 3, 227-241.
Moll, K. L. 1962. Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research* 5, 30-37.
Moll, K. L. and Daniloff, R. G. 1971. Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* 50, 678-684.
Ohala, J. J. 1971. Monitoring soft palate movements in speech. In *Project on Linguistic Analysis Reports* (Phonology Laboratory, Department of Linguistics, University of California, Berkeley), 13. J01-J015.
Ohala, J. J. 1986. Phonological evidence for top-down processing in speech perception. In Perkell, J. S. and Klatt, D. H. (eds.): *Invariance and Variability of Speech Processes.* Hillsdale, NJ: Erlbaum. 386-401.
Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. and Jackson, M. 1992. Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92, 3078-3096.
Skolnick, M. L., McCall, G. N. and Barnes, M. 1973. The sphincteric mechanism of velopharyngeal closure. *Cleft Palate Journal.* 10, 286-305.
Stevens, K. N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P. B. and David Jr., E. E. (eds.): *Human communication: A unified view.* New York: McGraw Hill. 51-66.
Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3-46.
Stevens, K. N. 1998. *Acoustic Phonetics.* Cambridge, MA: MIT Press.
Stevens, K. N. 2003. Acoustic and perceptual evidence for universal phonological features. *Proc. of International Congress of Phonetic Sciences.* 33-38.
Stevens, K. N., Fant, G. and Hawkins, S. 1986. Some acoustical and perceptual correlates of nasal vowels. In Channon, R., and Shokey, L. (eds.): *In Honor of Ilse Lehiste.* Dordrecht, Holland: Foris. 241-254.
Stevens, K. N., Nickerson, R. S., Boothroyd, A. and Rollins, A. M. 1976. Assessment of nasalization in the speech of deaf children. *Journal of Speech and Hearing Rearch* 19, 393-416.
Takeuchi, S., Kasuya, H. and Kido, K. 1977. A study on the effects of nasal and paranasal cavities on the spectra of nasal sounds. *Journal of the Acoustical Society of Japan* 33, 163-172 (in Japanese).
Ushijima, T. and Sawashima, M. 1972. Fiberscopic observation of velar movements during speech. *Annual Bulletin* 6. Tokyo: Research Institute of Logopedics and Phoniatrics, University of Tokyo. 25-38.
Wright, J. T. 1975. Effects of vowel nasalization on the perception of vowel height. In Ferguson, C. A., Hyman, L. M., and Ohala, J. J. (eds.): *Nasalfest: Papers from a Symposium on Nasals and Nasalization.* Stanford University, Stanford, CA: Language Universals Project. 373-388.
Wright, J. T. 1986. The behavior of nasalized vowels in the perceptual vowel space. In Ohala, J. J. and Jaeger, J. J. (eds.): *Experimental Phonology.* Orlando, FL: Academic Press. 45-67.

BEA – A MULTIFUNCTIONAL HUNGARIAN SPOKEN LANGUAGE DATABASE

Mária Gósy
Research Institute for Lingusitics, Hungarian Academy of Sciences
e-mail: gosy.maria@nytud.mta.hu

Abstract

In diverse areas of linguistics, the demand for studying actual language use is on the increase. The aim of developing a phonetically-based multi-purpose database of Hungarian spontaneous speech, dubbed BEA[2], is to accumulate a large amount of spontaneous speech of various types together with sentence repetition and reading. Presently, the recorded material of BEA amounts to 260 hours produced by 280 present-day Budapest speakers (ages between 20 and 90, 168 females and 112 males), providing also annotated materials for various types of research and practical applications.

1 Introduction

The creation of large speech databases with the help of computer technology has been called the "third revolution in the history of phonetics" in an opening address of a 2011 phonetics workshop at the University of Pennsylvania (http://www.ling.upenn.edu.phonetics/workshop/), the first two revolutions being the introduction of spectrographic analysis and that of computerized speech analysis software. Today, very large written and spoken databases are available in a number of languages; consequently, researchers can find answers to questions that, in the absence of relevant linguistic material, were simply unanswerable earlier on. A philological approach to texts does not have to be restricted to written corpora any more. In diverse areas of linguistics, the demand for studying actual language use is on the increase. Rule-based methods have been replaced by statistical ones in many cases as a result of the need to process very large quantities of data, a fact that has necessarily been accompanied by changes in researchers' attitudes, too.

Contemporary speech databases include structured sets of recordings and can be searched in a number of ways. Most of them are audio recordings but video recordings also exist (e.g. CUAVE: Patterson et al., 2002; Popescu-Belis et al., 2009). Nearly all databases include text files containing various levels of transcriptions of the recorded speech material. Databases can be classified in several ways – in terms of their respective aims, contents, written transcripts, circumstances

[2] The acronym BEA stands for the letters of the original name of the database: BEszélt nyelvi Adatbázis 'Speech Database'.

of recording, etc. (see e.g. Clark and Fox Tree, 2002). Some of them involve read texts, some contain spontaneous speech material, and some include both types of speech. Read materials usually involve parts of books, radio news items, word lists, etc. Spontaneous samples are recorded in laboratories, via telephone, or in field work, or else are selected from programs of the mass media; they may involve dialogues, conversations, narratives, real life situations (or their imitations), game situations, texts recorded using the map task method, etc. (see e.g. Anderson et al., 1991; Hennebert et al., 2000; Ruhi, 2011). Some of the large databases will be mentioned here. The British National Corpus is a collection including 100 million running words (Burnard and Aston, 1998). The London–Lund Corpus contains 50 dialogues and a mere 170,000 words (Svartvik, 1990). The (original) American English corpus CallHome includes 120 dialogues of 30 minutes on average, all of them family conversations via the phone. One of the earliest corpora was the Kiel Corpus, consisting of German spontaneous speech samples (Simpson et al., 1997). The HCRC Map Task is a database of mainly Scottish English speech involving 62 speakers and 18 hours of speech material using the map task (Anderson et al., 1991). The speech corpus of Stanford University called Switchboard (Godfrey et al., 1992; Calhoun et al., 2010) includes 2400 dialogues of 543 speakers (representing a number of American English dialects). TIMIT is used for training speaker-independent speech recognizers and includes 630 speakers reading 10 sentences each (Keating et al., 1994). The CSJ (Corpus of Spontaneous Japanese) contains 661 hours of speech by 1395 speakers including 7.2 million words (Maekawa, 2003). The Verbmobil database (Bael et al., 2007) has been developed with speech technological purposes in mind. Two databases representing seven European languages are EUROM1 and BABEL; their objective is to help the work of experts on speech acoustics, phonetics, digital signal processing and/or linguistics by providing recordings of various read texts (Chan et al., 1995; Vicsi, 2001).

As far as is known, a Hungarian database was first compiled by József Balassa at the beginning of the twentieth century; however, the material has been destroyed (see KKA 1994). In the 1940s, at the initiative of phonetician Lajos Hegedűs, dialect recordings started being made with the aim of recording speech, storytelling, magic formulae, etc. at various locations of the country; this material was archived in the late nineties on contemporary data carriers and made suitable for research in the Research Institute for Linguistics of the Hungarian Academy of Sciences (Gósy et al., 2011). The Budapest Sociolinguistic Interview (Budapesti Szociolingvisztikai Interjú, BUSZI) contains tape recorded interviews with 250 speakers (2–3 hours each) made in the late eighties (Váradi, 2003). The material has since been transcribed and encoded in computer files. The Hungarian telephone speech database MTBA is a speech corpus recorded via regular phone and cell phone in order to support Hungarian research and development in speech technology, containing read speech by 500 subjects (Vicsi et al., 2002; Vicsi, 2010). The HuComTech Multimodal Database contains audio-visual recodings (about 60 hours) of 121 young adult speakers that represent North-East Hungary (Pápay, 2011).

Speech databases usually contain some kind of written transcripts along with the recorded sound material. Depending on the area of utilization, such transcripts may be orthographic texts, phonemic (broad) transcriptions, or phonetic (narrow) transcriptions; they can include intonation and other suprasegmental details, etc. Along with individually developed systems, various kinds of universal and/or adaptable software are also available for providing transcripts (e.g. Praat: Boersma and Weenink, 2011; ToBI: Beckman et al., 2007). A complete system is offered by EXMARaLDA (Extensible Markup Language for Discourse Annotation: Schmidt 2009), specifically developed for the annotation of spoken language. The specifics of spoken language transcription, its degree of detail, form, and criteria may vary, depending on the aim or application involved (e.g. Grønnum, 2009; Maekawa, 2003). The fundamental difficulty of annotation resides in the fact that it is usually a single person (phonetician, linguist) who makes the transcription, hence the result unavoidably reflects, even if to diverse degrees, the transcriber's subjective perception (cf. Hunston, 2002). Transcription is a time consuming activity; its total duration includes a first listening to the given portion of text, several runs of repeated listening, preparing the written version of the given portion, its checking with repeated listening again, and correction (if any) by the primary transcriber or by another person.

The aim of the present paper is to provide an introduction to the development, results, and research possibilities of BEA, a spoken language database being developed at the Phonetics Department of the Research Institute for Linguistics of the Hungarian Academy of Sciences. This is the first Hungarian database of its kind in the sense that it involves many speakers, a very large amount of spontaneous speech material, with its transcripts and annotations of various levels, whose recording conditions are permanent and of studio quality. This structured collection of speech materials and their annotations makes directed search and the tabulation of data possible.

2 The development of the BEA database

Phonetic analyses in the strict sense, and linguistic analyses of a looser kind, that is, a multi-aspect research on spontaneous speech, made it necessary to develop a multifunctional database that can serve as a basis for both theoretical and applied studies. On the basis of experiences with existing corpora and databases, the development of BEA began in 2007. The long-term aim was recording speech from 500 speakers with gender and age proportions as well as level of schooling being represented in a balanced manner. In designing the contents (protocols) of the database we took the needs of the above research areas into consideration, we applied the most up-to-date recording techniques available when the data collection was started, and observed sociological factors to some extent (although this was not a primary consideration). At the same time, we started devising the transcription strategies and methods of data search to be made available. The long-term aim here is to provide a fully annotated and structured speech database. (In the development

of this database, the requirements of “Ethical Regulations of Experimental Research in Linguistics Involving Human Subjects” of the Research Institute for Linguistics of the Hungarian Academy of Sciences have been strictly observed in all respects.)

At the time of writing, the total recorded material of BEA is 260 hours, meaning approximately 3,300,000 running words. The shortest recording lasts 24 minutes and 27 seconds, the duration of the longest is 2 hours, 24 minutes and 47 seconds; the average length is 52 minutes. Two recordings are longer than 2 hours while 4 of them shorter than half an hour. There majority of them appear between 40 and 60 minutes (Figure 1).

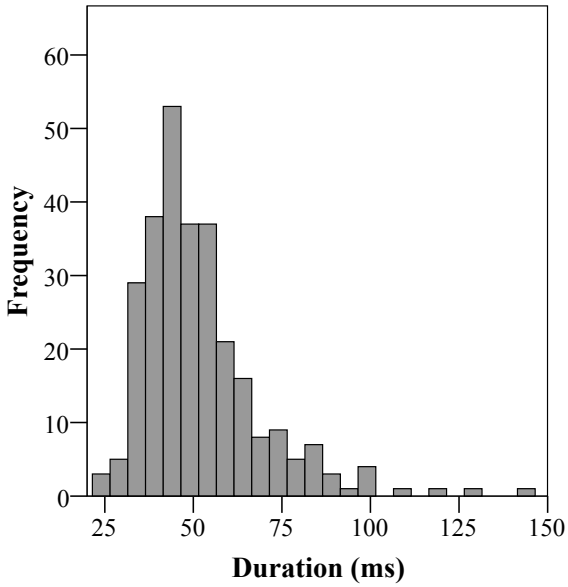


Figure 1. Various durations of recorded speech samples per speaker

2.1 The recording protocol of the BEA database

The database primarily contains spontaneous speech materials, but for the sake of comparisons, it also includes sentence repetitions and read texts. The protocol consists of six modules, labeled narrative, opinion, précis, conversation, sentence repetition, and reading. Various types of spontaneous speech are recorded with each of the subjects. 1. Narratives are about the subject’s life, family, job, and hobbies; they are more or less continuous monologues. 2. Opinions (that are mainly narratives, too) are requested about a topic of current interest, provided by the interviewer. The topics include getting one’s driver’s license, zero tolerance to the consumption of alcohol while driving, prospective price increases, marriage contract, climate change, violence against teachers, traffic in Budapest, home birth, online vs. traditional libraries, animal protection laws, small children’s use of cell phones, reading habits, mountains of debt, no smoking in public places, fat tax. The interviewer tries to make sure that the subject speaks fluently for as long as possible,

but this communicative situation requires that the interviewer also makes a point every now and then; hence dialogue-like situations may also arise. (The interviewer invariably tries to assume a standpoint that is opposite to that of the subject.) 3. Précis (summary of content) is in fact directed spontaneous speech. The subject hears a recorded text and then s/he has to summarize its content in his/her own words. One of the texts is a short item of popular science (174 words; 1 minute and 37 seconds), the other one is a funny story (270 words; 2 minutes and 5 seconds); both were recorded with an average female speaker. 4. In the conversation module, there are three participants: the subject, the interviewer, and a third person. The topics vary, but invariably concern everyday life; they have to differ from that of the opinion module of the same subject. Some conversation topics are: New Year's Eve, wedding experiences, job hunt, drug cultivation in one's home, Easter, marriage vs. cohabitation, secondary school final exams, summer holidays, preparations for Christmas, gas crisis in Europe, school violence, keeping pets in an apartment, the effect of economic crisis on culture, subway construction, legalization of light drugs, theatrical life, students' rights, women's careers, bringing up children, cycling as a form of traffic, concerts, the value of a university degree, etc. Topics for the opinion and conversation modules are selected by the interviewer in accordance with the subject's age, job, and area of interest (based on the narrative module). 5. The material for the sentence repetition module consists of 25 simple or compound sentences (e.g. *A farsangi bálban mindenkinek szép jelmeze volt* 'At the carnival dance, everyone wore nice fancy dresses'). The sentence is read out by the interviewer, and the subject has to repeat it immediately after a single hearing. (If the repetition is unsuccessful, the sentence may be read again by the interviewer.) 6. According to the protocol, the subject finally reads two texts aloud. One of them consists of the 25 sentences that the subject had to repeat earlier, the other one is an article taken from popular science.

2.2 Recording conditions

Recordings are invariably made in the same room, under identical technical conditions: in the sound-proof booth of the Phonetics Department, specially designed for the purpose. The size of the room (not counting the sound damping layer on the walls) is 340x210x300 cm. The degree of sound damping as compared to the outside environment is 35 dB at 50 Hz, and ≥ 65 dB above 250 Hz. The walls of the room are provided with a sound-absorbing layer in order to avoid reverberation. The way to the corridor is closed by double doors, with 30 cm distance in between; both doors can be opened and closed separately and are of a sound damping quality. The inner door has special noise insulation. The recording microphone is AT4040. Recording is made digitally, direct to the computer, with GoldWave sound editing software, with sampling at 44.1 kHz (storage: 16 bits, 86 kbytes/s, mono). The total size of recordings at present amounts to 71 GB; they are archived also on DVDs and on six external HDDs. In 95% of all recordings, the interviewer was the same young woman. The third participant of conversations was a young man or a young woman (researchers of the department).

3 Subjects

The number of subjects at present is 280; they are all monolingual adults from Budapest, not one of them reported any hearing disorders. At the moment, materials from 168 female and 112 male speakers are available. Their ages range between 20 and 90 years (Figure 2). In the future, as already noted, we will aim at a more balanced representation of age groups.

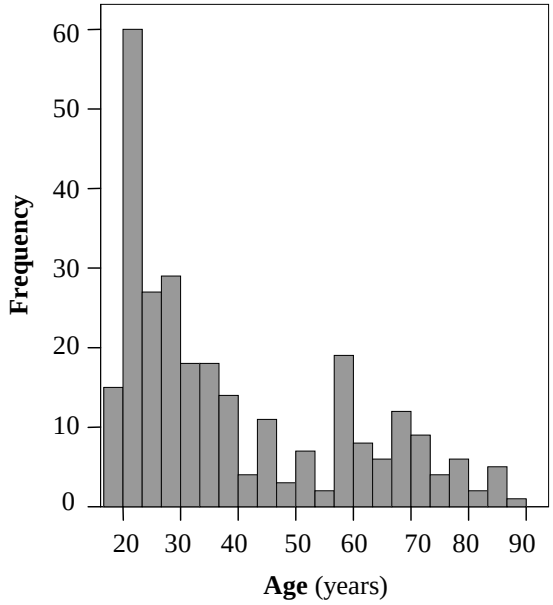


Figure 2. The distribution of BEA speakers by age

The recordings are anonymized (the speakers are given codes); they can be polled without identifying the given speaker. For each recording, the following data are documented: the subject's age, schooling, job, stature (height), weight, whether s/he is a smoker, and the topics of the spontaneous speech modules. Of the current group of subjects, 51 are smokers, 4 are ex-smokers (and 225 are non-smokers). 10 subjects completed 8th grade, 117 completed 12th grade (have taken secondary school final exams), 2 subjects have vocational degrees, and 161 have college/university degrees. Their jobs are extremely varied, including the following: district nurse, engineer, teacher, cleaner, teacher of children with disabilities, car mechanic, stoker, actor, office worker, paramedic, university student, media worker, payroll clerk, singer, housewife, organ builder, civil servant, tailor, physician, information specialist, store man, unemployed person, caretaker, economist, graphic artist, lifeguard, welder, delivery-man, priest, garden builder, poker player, scriptwriter, tile stove builder, nurse, game developer, real estate broker, etc.

The speaker's height and weight may be more or less closely related to his/her speech ('stature harmony', see Gósy, 1999). In some (applied) research and practical

applications (e.g. forensic phonetics), the estimability of weight and stature may be important (Figure 3).

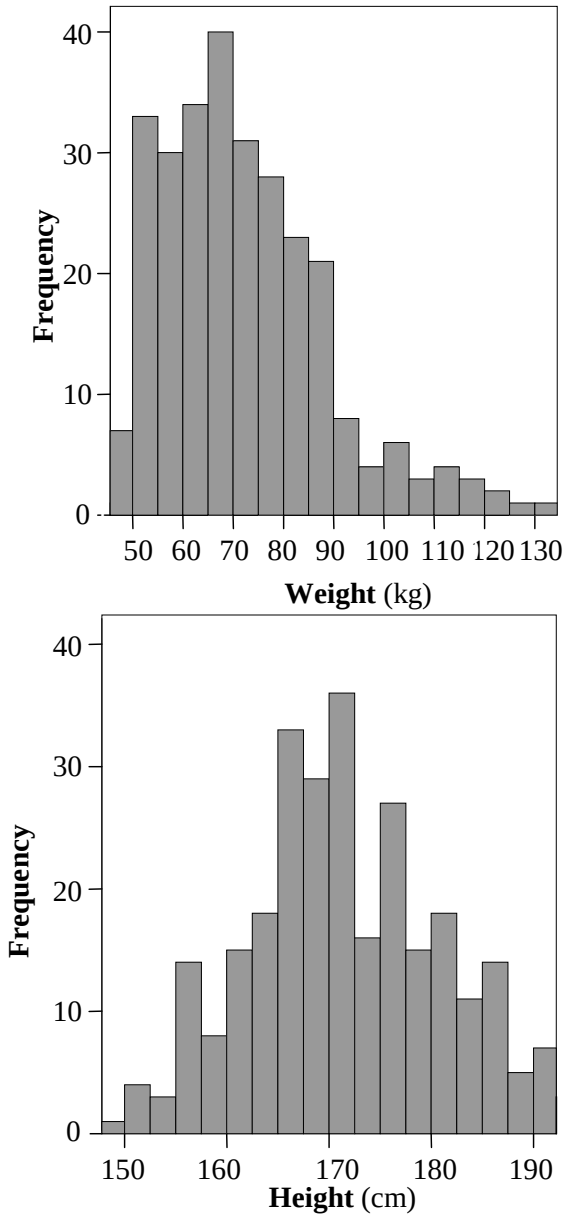


Figure 3. The distribution of BEA speakers by body weight and by stature

4 Transcription and annotation

The transcription of BEA materials is done at several levels. This fact makes it possible for the researcher to choose the most suitable level and use it in her/his

work. Transcripts of the various levels furthermore allow for gradualness from cursory overview to detailed annotation. At present, the following types of transcripts serve linguistic research and speech technology purposes.

1. Primary transcription in orthography but without punctuation. Transcribers use Microsoft Office Word (.doc format). The participants are uniformly abbreviated as A (subject), T1 (interviewer and first conversation partner), T2 (second conversation partner). According to the transcription regulations (see Gyarmathy and Neuberger, 2011), only proper names are capitalized, while phenomena that might be important in later phases are marked: disfluencies (bold), physiological and other nonverbal noises like laughter (exclamation mark), as well as speaking simultaneously (parentheses). The transcription uses emboldening for all nonstandard/erroneous forms; if the speaker does not add any correction, the transcriber adds the expected form in square brackets: *érzezzük* [érezzük] *magunkat* 'we have a grood [good] time'. Disfluency phenomena are uniformly marked: lengthening by doubling the given letter, hesitations (filled pauses) by triple letters (e.g. *öööö* 'er', *mmm* 'mmm'), and pauses, when perceived, by square marks (□) and non-verbal sounds by exclamation marks (!). All disfluency phenomena are written in bold letters. The transcription manual includes rules for transcribing words that occur as colloquially used but not in their dictionary form (e.g. *asszem* instead of *azt hiszem* 'I think'), foreignisms, abbreviations, acronyms, and forms that the transcriber finds unintelligible (enclosed between **) (see Figure 4). Transcriptions furthermore include duration data for the whole recording and for each module separately. Approximately 63% of the BEA database has so far been provided with accompanying primary transcripts.

az azt figyeltem meg hogy ! hogy akik **öööö** mondjuk így vezetgetnek **öööö** □ **öööö** egy-egy pohár alkohollal azok nem nagyon tudják megállni az egy **s** [sört] egy pohár sört hanem akkor betesznek mellé még két unikumot [unicumot] meg ! három pohár **öööö izé mmm** □ mit tudom én **mmm** □ királyvizet és **akkor** ! **akkor** az már nagyon erős

'I I noticed that ! that people who **er** say tend to drive their cars **er** □ **er** with a glass of alcohol or two they cannot easily stop with one **b** [beer] one glass of beer but they add two shots of Unicum and ! three glasses of **er whatsit mmm** □ how should I say **mmm** □ aqua regia and **then** ! **then** that is very strong indeed'

Figure 4. Sample fragment of conversation in primary transcription

Primary transcriptions have advantages and disadvantages. It is a good thing that the whole protocol can be included in a single file (per speaker), and thus words, word boundaries, nonverbal phenomena, etc. can easily be searched (automatically) in the transcript. What is not so good is that the transcript is difficult to synchronize with the sound material: it takes some time and some practice.

2. Annotation. This form of transcription is a kind of visual display of spoken texts and some further pieces of information related to them in a way that the written text and the actual recording can be displayed/listened to simultaneously. This is made possible by software like Praat and Transcriber. Praat is a complex acoustic signal processor, making annotation possible among other functions (Boersma and Weenink, 2011). The Transcriber program has been specifically developed for segmenting, labeling and transcribing spoken texts (see trans.sourceforge.net). Both programs have a user-friendly graphic interface and can use a number of platforms (Windows, Unix) (see Allwood et al., 2003; Weisser, 2003). As these are both English-language software programs, the controller interface (as well as the automatic labels in the case of Transcriber) appears in English. By default, transcribed texts can be stored and managed in .txt/TextGrid data files in Praat, and .trs files in Transcriber.

In Praat, phrases are defined as portions of speech between silent pauses (the latter identified by perceptual and visual information). In addition, turns (turn taking and turn yielding), background channel signals and the various types of pauses are also indicated. Transcription is primarily done in orthography without punctuation. Several types of annotation can be displayed in Praat (for an example see Figure 5).

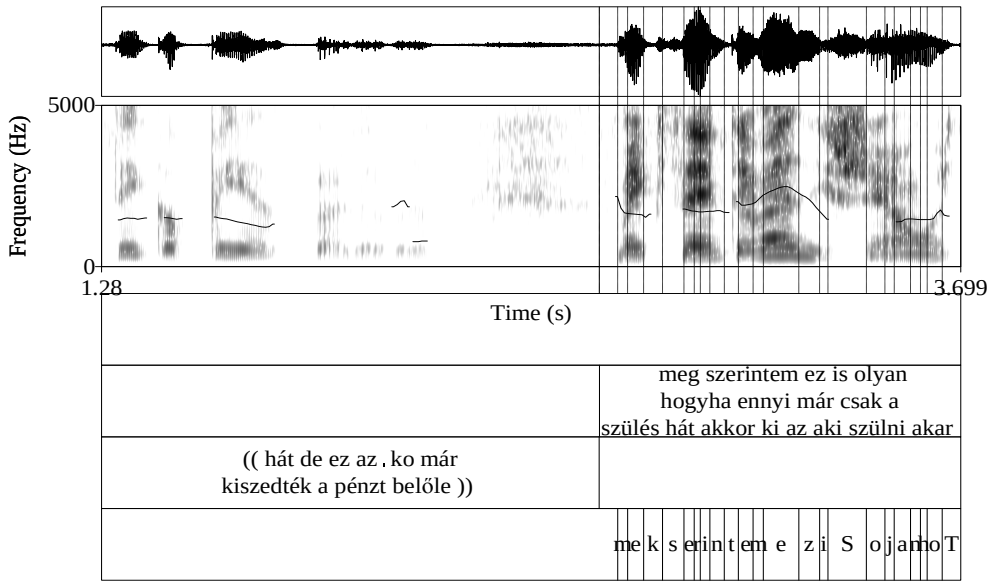


Figure 5. Sample annotation in Praat (the spectrogram shows the speech fragment ték a pénzt belőle meg szerintem ez is olyan hogy ‘took the money out and I think this is again so that’)

The vertical lines shown are segment boundaries. The sound level annotation occasionally uses capital letters (e.g. S stands for [ʃ]); this follows from the use of the automatic segmentation program (MAUS, cf. Beringer and Schiel 2000). Cases

of simultaneous speech, as well as unintelligible or hardly intelligible portions, are indicated by double parentheses. Some 10% of the recordings of the BEA database have been annotated so far in Praat; ten interviews are labeled at the phrase, word, and sound levels.

The Transcriber program allows for the segmentation, labeling and description of speech, especially for speech technology applications. The sound material and the written text can both be simultaneously made visible and audible here, too. The software supports several types of audio files (.au, .wav, .snd). Transcriber is also suitable for the automatic labeling of silent pauses, hesitations, as well as of other, non-speech vocalizations (e.g., coughing, laughing, and other noises) (Figure 6). Segmentation is done in terms of phrases, with their boundaries located at the middle of the silent pause between two phrases (the length of silent pauses is not shown but those thought to be longer than usual are indicated by the label [sil], cf. Gyarmathy and Neuberger, 2011). When the sound file is opened, the bottom of the display shows the oscillogram with single-level labeling below it (this is where vertical lines indicate segment boundaries) and a surface for typing in texts (indicating speakers, topics, etc.) above it (occupying most of the screen). In Transcriber, labeling is done in orthography; but in some cases (e.g. acronyms, foreign words, or old family names) pronunciation can be indicated, too. At present, 40% of all BEA recordings are annotated in Transcriber. There is about 5% overlap between the annotated speech samples of Praat and Transcriber.

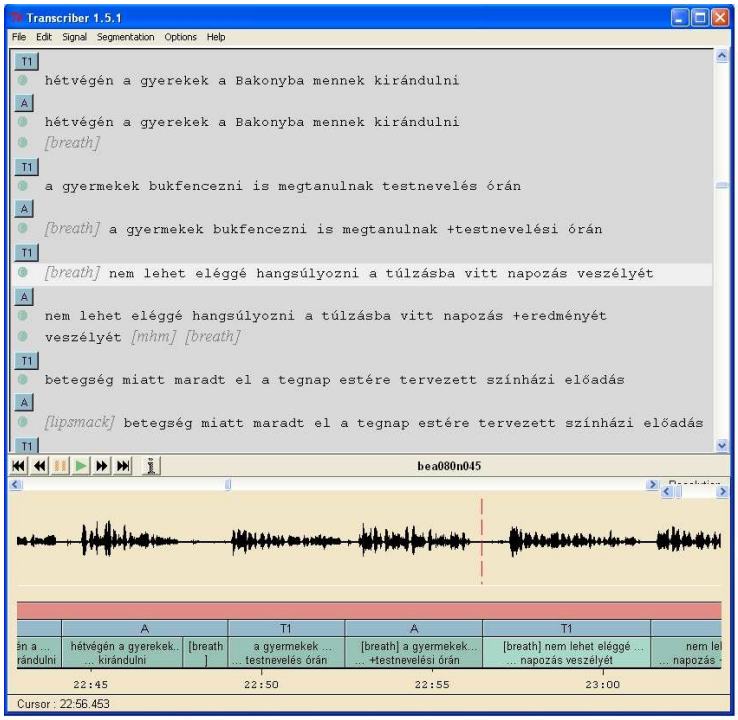


Figure 6. User's interface of Transcriber

References

Ahrens, B. 2005. Prosodic Phenomena in Simultaneous Interpreting: A Conceptual Approach and its Practical Application. *Interpreting* 7(1), 51-76.

Andrews, D. R. 1999. *Sociocultural perspectives on language change in diaspora: Soviet immigrants in the United States*. Amsterdam: John Benjamins.

Barik, H. C. 1972. Interpreters Talk a Lot, Among Other Things. *Babel* 18(1), 3-10.

Barik, H. C. 1973. Simultaneous Interpretation: Qualtitative and Linguistic Data. *Language and Speech* 16(3), 237-270.

Boersma, P. and Weenik, D. 1998. *Praat: doing phonetics by computer* (Version 5.0.1), http://www.fon.hum.uva.nl/praat/download_win.html.

Bóna, J. and Imre, A. 2007. A hangsúlyeltolódás hatása a beszédfeldolgozásra. [The effects of stress shift on speech perception.] *Beszédkutatás 2007.* 75-82.

Collins, B. and Mees, I. M. 2008. *Practical Phonetics and Phonology*. London: Routledge.

Cutler, A. 1980. Errors of stress and intonation. In Fromkin, V. A. (ed.): *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York–London: Academic Press. 67-80.

Cutler, A. 2008. Lexical Stress. In Pisoni, D. B. and Remez, R. E. (eds.): *The Handbook of Speech Perception.* Malden, MA–Oxford: Blackwell Publishing. 264-289.

Cutler, A., Dahan, D. and van Donselaar, W. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40, 141-201.

Cutler, A. and Ladd, R. D. (eds.) 1983. *Prosody: Models and measurements.* Berlin–Heidelberg–New York–Tokyo: Springer.

Fónagy, I. 1958. *A hangsúlyról.* [On stress.] Budapest: Akadémiai Kiadó.

Goldman-Eisler, F. 1972. Segmentation of Input in Simultaneous Translation. *Journal of Psycholingusitic Research* 1(2), 127-140.

Gile, D. 1995. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam, Philadelphia: John Benjamins.

Gósy, M. 2002. A hangsúlyeltolódás jelensége. [The phenomenon of stress shift.] In Balázs, G., A. Jászó, A. and Koltói, Á. (eds.): *Éltető anyanyelvünk.* Budapest: Tinta Könyvkiadó. 193-198.

Gósy, M. 2004. *Fonetika, a beszéd tudománya.* [Phonetics] Budapest: Osiris Kiadó.

Hardcastle, W. J. and Laver, J. (eds.) 1999. *The Handbook of Phonetic Sciences.* Oxford: Blackwell.

Kálmán, L. and Nádasdy, Á. 1994. A hangsúly. [The stress] In Kiefer, F. (szerk.): *Strukturális magyar nyelvtan 2. Fonológia* [A Structural Grammar of Hungaran 2. Phonology]. Budapest: Akadémiai Kiadó. 393-468.

Keszler, B. (ed.) 2000. *Magyar grammatika* [Hungarian grammar.] Budapest: Nemzeti Tankönyvkiadó.

Laver, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lee, T.-H. 1999. Speech proportion and accuracy in simultaneous interpretation from English into Korean. *Meta* 44(2), 260-267.

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation.* A Bradford Book. Cambridge (Massachusetts)–London (England): The MIT Press.

Mennen, I. 2004. Bidirectional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32, 543-563.

Nooteboom, S. 1999. The prosody of speech: Melody and rhythm. In Hardcastle, W. J. and Laver, J. (eds.): *The Handbook of Phonetic Sciences.* Oxford: Blackwell. 640-674.

Olaszy, G. 2002. Predicting Hungarian sound durations for continuous speech. *Acta linguistica Hungarica* 49(3-4) 321-345.

Paradis, M. 2000. Prerequisites to a Study of Neurolinguistic Processes involved in Simultaneous Interpreting. A Synopsis. In Dimitrova, E. and Hyltenstam, K. (eds.): *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives.* Amsterdam, Philadelphia: John Benjamins. 17-24.

Roach, P. 1992. *English Phonetics and Phonology.* Cambridge: Cambridge University Press.
Romaine, S. 1989. *Bilingualism.* Oxford: Blackwell Publishing.
Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica* 23, 1-33.
Shlesinger, M. 1994. Intonation in the production and perception of simultaneous interpretation. In Lambert, S. and Moser-Mercer, B. (eds.): *Bridging the gap: Empirical research in simultaneous interpretation.* Amsterdam: John Benjamins. 225-236.
Spiller, E. and Bosatra, A. 1989. Role of the Auditory Sensory Modality in Simultaneous Interpretation. In Gran, L. and Dodds, J. (eds.): *The Theoretical and Practical Aspects of Teaching Conference Interpretation.* Udine: Campanotto Editore. 37-38.
Szende, T. 1995. *A beszéd hangszerelése. Idő, hangmagasság, hangerő és határjelzés a közlésben* [Time, pitch, volume and boundary marking in utterances]. Budapest: MTA Nyelvtudományi Intézet.
Toury, G. 1995. *Descriptive Translation Studies and beyond.* Amsterdam: John Benjamins.
Vaissière, J. 2008. Perception of Intonation. In Pisoni, D. B. and Remez R. E. (eds): *The handbook of Speech Perception.* Malden, MA–Oxford: Blackwell Publishing. 236-263.
Varga, L. 1985. Intonation in the Hungarian sentence. In Kenesei, I. (ed.): *Approaches to Hungarian. Volume one. Data and descriptions.* Szeged: JATE. 205-224.
Varga, L. 2000. A magyar mellékhangsúly fonológiai státusáról. *Magyar Nyelvőr* 124, 91-108.
Williams, S. 1995. Observations on anomalous stress in interpreting. *The Translator* 1(1), 47-64.

RESEARCH DEPARTMENT OF SPEECH, HEARING AND PHONETIC SCIENCES, UCL

http://www.ucl.ac.uk/psychlangsci/research/speech

1 The department and its location

Phonetics and Speech Science at University College London (UCL) is based in the Research Department of Speech, Hearing and Phonetic Sciences (SHaPS) (http://www.ucl.ac.uk/psychlangsci/research/speech). This group is one of the largest in the field in the UK, and is part of a thriving research community with close links to many other departments and institutes within UCL, including particularly the UCL Ear Institute and the Institute of Cognitive Neuroscience. At the time of writing the department has 8 academic staff and 7 post-doctoral researchers. The department is part of the Division of Psychology and Language Sciences, which brings together researchers in a range of disciplines such as cognition, neuroscience, education, communication, medicine, and health, as well as phonetics and linguistics, within UCL's Faculty of Brain Sciences.

The present grouping was created in 2008 when the former department of Phonetics and Linguistics was relocated alongside the former department of Human Communication in Chandler House, about 1 km east from the main UCL campus. The Victorian building, once the site of the Royal Free Hospital Medical School for Women, was refurbished with an investment of £13M to house state-of-art research and teaching facilities for Language Sciences on all five floors, including the Language and Speech Sciences Library (LaSS), which is both a branch of the main UCL library and also the National Information Centre for Speech-language Therapy (NICeST), holding a unique specialist collection of materials in the field of human communication and its disorders, covering language and languages (written, spoken and signed), linguistics, phonetics, psychology, special education, speech science and voice.

Teaching facilities within Chandler House include two lecture theatres, each with a capacity of 90, three large teaching rooms seating 40 to 60, a 12-workstation computer cluster room, eight smaller meeting or seminar rooms, five rooms of a working speech-language therapy clinic, and a Teaching Laboratory suitable for groups of up to 40 students, with 18 computer workstations.

2 Research facilities and equipment

A dedicated ground floor laboratory is available for work with children, forming the UCL Infant and Child Language (ICL) Research Centre, while the basement houses the main Research Laboratory for phonetics and speech science, upon which approaching half a million pounds was invested during the refurbishment. It provides seven double-walled air-conditioned listening/recording rooms with ambient noise levels below the threshold of hearing. All are large enough to

accommodate 2-3 people, one being rather larger and equipped for audio-visual recording. One room is additionally constructed as an electrically and magnetically shielded Faraday cage to allow the measurement of very low level electrical signals reflecting neural activity. The seven rooms are reached from a lobby which houses ancillary equipment and services, and equipment storage and workshop facilities are nearby, together with a kitchen, and a waiting area for subjects. There are patch panels to each of the rooms for audio, video, computer network and other data; a dedicated server is located in the building.

Figure 1. In the lobby of the Research Laboratory, giving access to the seven sound-treated listening/recording rooms

High-quality microphones are available for audio recording, generally accomplished by direct digital capture to the networked computers. The shielded room houses a 64-electrode EEG (Electroencephalography) and ABR (Auditory Brainstem Response) facility installed in 2012. Several UCL-developed Laryngographs® are available for electro-glottographic recording (commonly made as a second channel along with speech), while the Teaching Laboratory separately has some 15 Laryngograph units. Though the main focus of work has recently been on speech perception rather than production, the laboratory is also equipped for aerodynamic studies, using Rothenberg masks and transducers from Glottal Enterprises, and for accelerometric studies of nasality.

Figure 2. Inside one of the recording rooms. The experimenter is using a Rothenberg mask to gather airflow data. On the bench are a portable Laryngograph, calibration equipment for the airflow unit, and a PC adapted for multi-channel acquisition.

In addition to the Research Laboratory at Chandler House, the department retains the use of an Anechoic Chamber adjacent to the prior location of the department in Gordon Square. This was originally completed in 1948, and has subsequently been refurbished to even higher standards of performance, providing a recording environment with very low ambient noise and reflected sound. The ambient sound pressure is below the threshold of hearing, and reverberation is controlled so that free-field conditions exist above 90Hz. The chamber is equipped with a Bruel and Kjaer 2231 Sound Level Meter and various recording systems. Signals are routed to an adjacent control room. This facility is made available for hire by outside agencies.

3 Students and courses

The Research Department of Speech, Hearing and Phonetic Sciences currently has 16 PhD students, and a dedicated SHaPS PhD work room with individual workstations is provided on the third floor; research students attached to the other Language Sciences research departments have similar facilities elsewhere in the building. The speech research facilities and teaching are also relevant to a range of postgraduate and undergraduate programmes. Courses include an MSc in Language Sciences with specialisation in Speech and Hearing Sciences, MRes in Speech, Language and Cognition, a long-established MA Phonetics, and the clinical MSc in Speech and Language Sciences, offering training in speech and language pathology and therapy. Phonetics is also an important component of the BA Linguistics. Almost all Language Sciences teaching takes place in Chandler House.

4 Research projects

Current SHaPS research foci are exemplified by the appended list of selected publications from 2011 and 2012. They include speech perception by learners of English (Iverson, Hazan) and by users of cochlear implants (Faulkner, Rosen); voice

transformation in therapy for schizophrenia (Huckvale); the adaptability of talkers to noisy and distorting communication channels (Hazan); accent change in English (Evans); models of the production and perception of intonation and tone (Xu). The department is a partner in a Marie Curie training network INSPIRE (2012-2016) that supports PhD and post-doctoral training addressing the perception of speech in non-optimal environments. Other funding comes from the Medical Research Council, the Economic and Social Science Research Council, the UK Home Office, and from three charities, the Wellcome Trust, Action on Hearing Loss and Deafness Research UK.

5 Publications

A complete listing of research publications from present members of the research department, covering more than 30 years, is available at:
http://www.ucl.ac.uk/psychlangsci/research/speech/Research-Publications.

A number of book-length treatments have been produced over the years by members of the department. Two of these are widely adopted teaching texts at present: *Signals and Systems* by Rosen and Howell (second edition 2011) and *Introducing Phonetic Science* by Ashby and Maidment (2005).

For much of the twentieth century, the IPA was sustained by an executive and administrative hub at UCL, where its journal Le *Maître phonétique* (later *Journal of the International Phonetic Association*, JIPA) and other publications, such as the chart of the alphabet itself, were produced and distributed. The journal *Language and Speech* was started in the department in 1958; both that journal and JIPA are now refereed journals produced by commercial publishers.

6 Software and resources

The department makes available the Speech Filing System, a free computing environment for PCs for conducting research into the nature of speech. It comprises software tools, file and data formats, subroutine libraries, graphics, special programming languages and tutorial documentation. It performs standard operations such as acquisition, replay, display and labelling, spectrographic and formant analysis and fundamental frequency estimation. It comes with a large body of ready made tools for signal processing, synthesis and recognition, as well as support for custom software development. SFS began as a tool created for a large UK collaborative research project in 1987 (Alvey). SFS was initiated by by Mark Huckvale at UCL and has been maintained and developed by him continuously since that time. The department has for many years made available information on the use of phonetic symbols in computers, including the development of the SAMPA alphabet and the distribution of special fonts. With the almost universal use of Unicode, the department provides a freely-downloadable Unicode keyboard for Windows. SFS and many other resources can be obtained from http://www.ucl.ac.uk/psychlangsci/research/speech/resources, while recorded material

from the department, such as a CD of “Sounds of the IPA”, is sold via an online shop (http://www.phon.ucl.ac.uk/shop/).

The department also hosts and maintains the webpages of the International Phonetic Association (http://www.langsci.ucl.ac.uk/ipa/) and the proceedings of PTLC (Phonetics Teaching and Learning Conference: http://www.phon.ucl.ac.uk/ptlc/) a biennial international conference which takes place at UCL.

7 Origin of the modern laboratory

From 1972 until its move to Chandler House in 2008, the Phonetics/Speech Science Laboratory was primarily in Wolfson House, a newly-built annex a little to the north of the main UCL campus. This was a period of intense development of the department’s experimental facilities; the two-room laboratory and anechoic room in the department’s original home at 21 Gordon Square were now extended by a purpose-designed air-conditioned ensemble comprising four doubly-isolated recording rooms, ten listening booths, a large laboratory equipped with in-house designed speech science teaching equipment, three computer/experimental rooms, a workshop, small library, seven staff office rooms, a small kitchen, a Common Room with kitchen area and two rooms for secretaries, all on the ground floor.The basement housed a large lecture theatre and a dedicated store-room. Funding came from the Wolfson Foundation, the Department of Health, and from UCL investment in the Phonetics and Linguistics Department’s initiative in the introduction of a new clinical BSc Speech Sciences degree, BSc Speech Communication, and the MSc in Speech and Hearing Sciences and also from a series of successful external grant applications.

The laboratory was also very active in research and was, for example: a pioneer with Cambridge University and Guy’s Hospital in electro-cochlear stimulation research with MRC programme grant support. A major contributor to the UK Alvey Spoken Language Engineering initiative in the mid 1980s; and the Coordinator of the EU flagship Speech Assessment Methods (SAM) project involving 26 Phonetics and Speech Science laboratories in eight European countries. This put the department to the fore in Europe and fostered considerable international collaboration and exchange. Research for some twenty successful PhD theses was completed, leading to the definition of a number of current research themes. The holders of some current senior research and teaching posts joined over this time: Stuart Rosen, Professor of Speech and Hearing Science, joined the department in 1977, initially to work on the cochlear implant project, Valerie Hazan, Professor in Speech Sciences and formerly Head of Department, came in 1980, and worked in the SAM consortium; Andrew Faulkner, currently Head of Department, joined in 1989, worked on and then led a series of UK and EU projects on hearing aids to support lipreading, while Mark Huckvale pioneered the Speech Pattern Audiometry research and initiated the MSc in Hearing and Speech Sciences.

From 1961 to 1972 the department’s laboratory facilities were in Gordon Square and this period saw: the first modernisation of the anechoic room, and the

construction, above it, of a speech science laboratory pneumatically isolated to curb the transmission of vibration; the introduction of the first flexible speech formant synthesis control system, copied by KTH (Kungliga Tekniska Högskolan) in Sweden and Bell Laboratories in the USA and used by government and university laboratories elsewhere in the UK (see Figure 6); and, for example, the first irrefutable experimental demonstration of the existence of central neuro-temporal pitch processing (Fourcin 1970); the invention and phonetic application of the Laryngograph (now in worldwide application) for language research, deaf voice training, voice therapy/pathology and the associated introduction of quantitative voice analysis based on connected speech (Fourcin and Abberton, 1971).

From 1961 until 1992, Adrian Fourcin, Professor of Experimental Phonetics, was the Head of the Phonetics/Speech Sciences Laboratory. An interview in which he describes aspects of his training and early career can be read here: http://americanhistory.si.edu/archives/speechsynthesis/ss_four.htm.

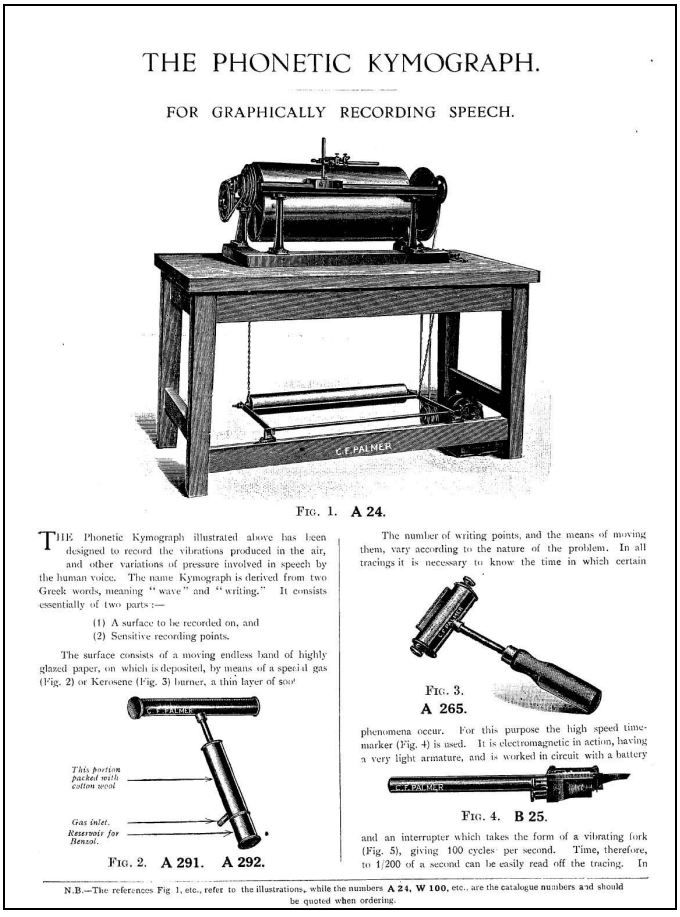
Figure 3. For more than 40 years, acoustics teaching has included hands-on lab practicals for every participant, with specially-developed apparatus. Here Speech Sciences students are introduced to the essential nature of formants by measuring the characteristics of an acoustic resonator. After confirming that the same laws govern mechanical, acoustic and electrical resonators, students are led to understand the use of electrical resonators in signal analysis.

Figure 4. Interactive speech perception testing in the mid-1990s at Wolfson House (Valerie Hazan and subject), using formant synthesis and a touch sensitive response box—before the advent of touch sensitive computer screens. Two spectrographs can be glimpsed in the background, on the left Houde's real-time and on the right the drum of a Kay mechanical scanner.

8 History: the first fifty years

The history of phonetics at UCL extends back more than a century. Daniel Jones (1881-1964), who was to become Britain's first professor of phonetics, began

lecturing at UCL in 1907. His career, and the development of the department he created, are documented in Collins and Mees (1998). Experimental work began from around 1912 with the appointment of Stephen Jones (1872-1942) who became the first superintendent of the laboratory (the two Joneses were unrelated). The laboratory techniques of his day included static X-ray photography (Jones, S., 1929), indirect palatography, the use of sensitive flames as sound detectors, and the measurement of voicing and duration by means of the kymograph. Stephen Jones supervised the construction of a kymograph with an unusually large electrically-driven drum, facilitating accurate measurements of speech sound duration and fundamental frequency, and the design was put into production by the firm of C. F. Palmer, and purchased for installation in other laboratories around the world (Figure 5). Film from 1928, showing Stephen Jones operating a kymograph of this type, has recently been discovered and restored (Ashby, 2011), and can be seen at: http://youtu.be/cXp7jfgRNVA.



THE PHONETIC KYMOGRAPH.

FOR GRAPHICALLY RECORDING SPEECH.

C.F.PALMER

FIG. 1. **A 24.**

THE Phonetic Kymograph illustrated above has been designed to record the vibrations produced in the air, and other variations of pressure involved in speech by the human voice. The name Kymograph is derived from two Greek words, meaning "wave" and "writing." It consists essentially of two parts:—

(1) A surface to be recorded on, and
(2) Sensitive recording points.

The surface consists of a moving endless band of highly glazed paper, on which is deposited, by means of a special gas (Fig. 2) or Kerosene (Fig. 3) burner, a thin layer of soot

C.F.PALMER

This portion packed with cotton wool

Gas inlet. Reservoir for Benzol.

FIG. 2. **A 291. A 292.**

The number of writing points, and the means of moving them, vary according to the nature of the problem. In all tracings it is necessary to know the time in which certain

FIG. 3. **A 265.**

phenomena occur. For this purpose the high speed time-marker (Fig. 4) is used. It is electromagnetic in action, having a very light armature, and is worked in circuit with a battery

C.F.PALMER

FIG. 4. **B 25.**

and an interrupter which takes the form of a vibrating fork (Fig. 5), giving 100 cycles per second. Time, therefore, to 1/200 of a second can be easily read off the tracing. In

N.B.—The references Fig 1, etc., refer to the illustrations, while the numbers **A 24, W 100,** etc., are the catalogue numbers and should be quoted when ordering.

Figure 5. A page from the 1932 catalogue of C.F. Palmer, Ltd., showing the large horizontal kymograph developed by Stephen Jones, and some of its accessories.

UCL hosted the second International Congress of Phonetic Sciences in 1935 (Jones and Fry 1936), and though the department had by this stage become pre-eminent in the world, the basement laboratory remained relatively modest, serving a subordinate role to linguistic phonetic investigations.

Stephen Jones was succeeded as superintendent of the laboratory in 1937 by D. B. Fry (1907-1983) who additionally became Head of Department in 1949 when Daniel Jones retired, and Professor of Experimental Phonetics in 1958. He is best known for his widely cited experimental work on the perception of stress in English words, which showed that duration and pitch are much more powerful cues to stress than loudness. Fry recruited an engineer, Peter B. Denes (1920-1996), who worked in the department over the period 1946-1961, to assist in the energetic postwar expansion of experimental facilities and in the teaching of experimental phonetics. Fry and Denes worked together on the design and construction of a speech recognizer realized in analogue hardware, which had an unlimited vocabulary and incorporated "linguistic knowledge" in the form of phoneme transition probabilities. It can be seen in operation in a film which was shown at the fourth ICPhS in Helsinki in 1961, now available at: http://youtu.be/9lKf3Dm_pJA

Both Fry and Denes were effective teachers, and both produced successful foundation-level textbooks drawing on their experience. Denes was the co-author (with Pinson) of the popular book *The Speech Chain* published in 1963, shortly after he left UCL for Bell Laboratories, while after his retirement in 1975, Fry found time to write *The Physics of Speech* (1979).

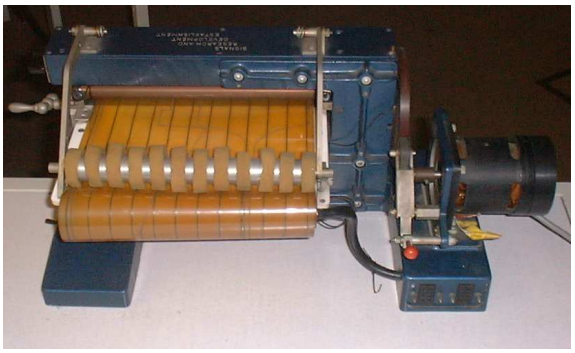


Figure 6. Hardware from the 1960s. In the interregnum between mechanical devices and computers, hybrid electronic equipment was used for experiments. The multi-function potentiometric speech synthesiser controller—familiarly known as the "mangle"—employed conducting ink tracks drawn on a flexible printed circuit belt (Fourcin 1960). It was devised by A. J. Fourcin and developed collaboratively at UCL and the Signals Research and Development Establishment (SRDE). This example, one of nine made, is now in the Smithsonian Museum in Washington DC; another is in the Science Museum in London.

9 Links with linguistic phonetics and practical training

Through most of its history the laboratory was a harmonious part of a department, which also placed great emphasis on linguistic phonetic description and on practical training in sound recognition and production. Many of the department's most celebrated members, such as Daniel Jones himself, Harold Palmer, J. R. Firth, A. C. Gimson, J. D. O'Connor and J. C. Wells, were not primarily experimentalists, though all were ready to accord a place for measurement alongside the findings of the trained ear. It was also taken for granted that those working in the laboratory should have a good practical training and be thoroughly familiar with phonetic symbols and classification. Practical phonetic training remains a vital component of such programmes as the (clinical) MSc in Language Sciences, and the MA Phonetics. A number of pedagogical initiatives have sought to bring the phonetics laboratory into the practical phonetics classroom, reinforcing the training of skill with immediate acoustic analysis of teachers' and students' productions (Ashby 2008).

References

Ashby, M. 2008. New Directions in Learning, Teaching and Assessment for Phonetics. *Estudios De Fonética Experimental* XVII, 19–44.

Ashby, M. 2011. Film from a Phonetics Laboratory of the 1920s. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 168–171. Hong Kong.

Ashby, M., and Maidment, J. 2005. *Introducing Phonetic Science*. Cambridge Introductions to Language and Linguistics. Cambridge: Cambridge University Press.

Collins, B. and Mees, I. 1998. *The Real Professor Higgins: The Life and Career of Daniel Jones*. Berlin: Mouton de Gruyter.

Denes, P. and Pinson, E. 1993. *The Speech Chain: The Physics and Biology of Spoken Language*. New York, N.Y: W.H. Freeman.

Jones, D. and Fry, D. 1936. *Proceedings of the Second International Congress of Phonetic Sciences: Held at University College, London, 22-26 July 1935*. Cambridge□: Cambridge University Press.

Jones, S. 1929. Radiography and Pronunciation. *British Journal of Radiology* 2, 149–150.

Fourcin, A. 1960. A potential dividing function generator for the control of speech synthesis. *Journal of the Acoustical Society of America* 32 (11), 1501.

Fourcin, A. 1970 Central Pitch and Auditory Lateralization. In Plomp, R., and Smoorenburg, G.F. (eds.) *Frequency Analysis and Periodicity Detection in Hearing*. Leiden: A.W. Sijthoff. Available at: http://discovery.ucl.ac.uk/1330855/1/Fourcin%20Central%20pitch%20final.pdf

Fourcin, A. and Abberton, E. 1971. First applications of a new laryngograph. *Medical and Biological Illustration* 21, 172-182. Reprinted (1972) *Volta Review* 69, 507-518.

Fry, D. 1979. *The Physics of Speech*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Rosen, S. and Howell, P. 2011. *Signals and Systems for Speech and Hearing*. 2nd ed. Bingley: Emerald.

Selected publications 2011/12

Granlund, S., Hazan, V. and Baker, R. 2012. An acoustic-phonetic comparison of the clear speaking styles of late Finnish-English bilinguals. *Journal of Phonetics* 40, 509-520.

Green, T., Faulkner, A. and Rosen, S. 2012. Variations in carrier pulse rate and the perception of amplitude modulation in cochlear implant users. *Ear and Hearing* 33, 221-230.

Hazan, V. and Baker, R. 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America* 130, 2139-2152.

Hilkhuysen, G., Gaubitch, N., Brookes, M. and Huckvale, M. 2012. Effects of noise suppression on intelligibility: dependency on signal-to-noise ratios. *Journal of the Acoustical Society of America* 131, 531-539.

Iverson, P., Wagner, A., Pinet, M. and Rosen, S. 2011. Cross-language specialization in phonetic processing: English and Hindi perception of /w/-/v/ speech and nonspeech. *Journal of the Acoustical Society of America* 130, EL297-EL303.

Lehtonen, M., Hultén, A., Cunillera, T., Rodriguez-Fornells, A., Tuomainen, J. and Laine, M. 2012. Differences in word recognition between early bilinguals and monolinguals: behavioral and ERP evidence. *Neuropsychologia* 50, 1362-1371.

Liu, F., Xu, Y., Patel, A. D., Francart, T. and Jiang, C. 2012. Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition* 79, 209-215.

McGettigan, C., Evans, S., Rosen, S., Agnew, Z., Shah, P. and Scott, S. 2012. An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience* 24, 636-652.

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H. and Scott, S. 2012. Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia* 50, 762-776.

Messaoud-Galusi, S., Hazan, V. and Rosen, S. 2011. Investigating speech perception in children with dyslexia: is there evidence of a consistent deficit in individuals? *Journal of Speech, Hearing and Language Research* 54, 1682-1701.

Pereira, V. J., Tuomainen, J. and Sell, D. 2011. The impact of maxillary osteotomy on speech outcomes in cleft lip and palate: an evidence-based approach to evaluating the literature. *The Cleft Palate-Craniofacial Journa,l* in press, http://dx.doi.org/10.1597/11-116.

Pinet, M., Iverson, P. and Huckvale, M. 2011. Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *Journal of the Acoustical Society of America* 130, 1653-1662.

Prom-on, S., Liu, F. and Xu, Y. 2012. Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *Journal of the Acoustical Society of America* 132, 421-432.

Rosen, S., Wise, R., Chadha, S., Conway, E. and Scott, S. 2011. Hemispheric asymmetries in speech perception: Sense, nonsense and modulations. *PLoS One* 6, e24672.

van Dommelen, W. and Hazan, V. 2012. Impact of talker variability on word recognition in non-native listeners. *Journal of the Acoustical Society of America* 132, 1690-1699.

Wang, B. and Xu, Y. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 39, 595-611.

Xu, Y., Chen, S-W., and Wang, B. 2012. Prosodic focus with and without post-focus compression PFC: A typological divide within the same language family? *The Linguistic Review* 29, 131-147.

Michael Ashby, Andrew Faulkner, Adrian Fourcin
Speech, Hearing and Phonetic Sciences, UCL
London, UK
e-mail: {m.ashby|a.faulkner|a.fourcin}@ ucl.ac.uk

ON THE HUNGARIAN SUNG VOWELS

Andrea Deme
Department of Phonetics, Eötvös Loránd University,
Research Institute for Linguistics, Hungarian Academy of Sciences
e-mail: deme.andrea@nytud.mta.hu

Abstract

Singing at a very high pitch is associated with vocal tract adjustments in professional western operatic singing. However, as of yet there is an inadequate amount of data available on the extent of the acoustic transformation the Hungarian vowels undergo during singing. The author's purpose is to evaluate the acoustic and articulatory changes of Hungarian vowel qualities, and examine the effect of these changes on the intelligibility of sounds, which has not yet been done for Hungarian. The paper contains a brief summary of formerly described tendencies for other languages and data for Hungarian from pilot studies carried out by the author with an adult soprano's and a child's sung vowels.

1 Theoretical introduction and questions

High-pitched singing in the western operatic style demands special articulatory movements and therefore is a specific object of analysis. Possible vocal tract adjustments one uses while singing have already been extensively described, but the effect of these modifications on the acoustic domain and the perception of sounds can differ considerably from language to language depending on the vowel system. Moreover, characteristics and registers of the high vocal range are less studied, because of its dependence on different techniques and training methods. Thus the aim of the research reported here is to investigate the effect of the articulation of singing on the Hungarian vowels /ɔ, aː, ɛ, eː, iː, oː, øː, uː, yː/ on production and perception as well. Assuming that singing and speech can be understood with equal ease, the operatic tradition does not have the practice of subtitling the performances played in the language of the audience (Watson, 2009). Consequently, our research by proving increased difficulty in percieving the high-pitched sung language elements, might point out the necessary change of this practice.

There is some agreement, that in speech, the vowel can be characterised by its first two formants (F_1, F_2) (Peterson and Barney, 1952), and these can also be a cue to their perception (Gósy, 1987; Neary, 1989). These resonances in adult speech normally lie far above the speaker's fundamental frequency (f_0), but in high-pitched singing, the f_0 is often raised above the average value of F_1 (or occasionally even the F_2). In this case, maintaining the normal vowel-dependent values of F_1 would not only change the timbre required for the western operatic style, but the singer should use greater vocal effort (along with unhealthy phonation) to provide the necessary loudness as well. To avoid producing weak sound (i.e. losing timbre and loudness),

when f_0 exceeds F_1 trained singers starts to tune F_1 (i.e. adjust the first resonance of the vocal tract) to the value of the raised fundamental (Sundberg, 1989; Garnier et al., 2010), therefore enhancing the amplitude of f_0. At high pitch this tuning can be described linearly (where F_1:f_0 tuning is controlled by jaw movement indirectly), but in lower ranges it can only be described by more complex nonlinear coupling effects (i.e. impedance-matching) (see Titze and Worley, 2009).

In general at high pitch, F_1 can be tuned by lowering the jaw and unrounding the lips (Sundberg, 1969). The position of the larynx is also changed with ascending pitch, but the direction and the nature of its movement seem to have great variability across singers. According to Sundberg (1969), the vertical position of the larynx in singing is (broadly speaking) inversely proportional to pitch. Hurme and Sonninen (1995) later described four basic movement strategies that can be observed in female and male singers: the larynx can be pulled in an 1) anterior-superior (up and forward), 2) posterior-superior (up and backward), or an 3) inferior (down) direction, but it also can have a 4) complex, zig-zagging route while raising f_0. Hurme and Sonninen also showed that the cartilages and the hyoid bone can change their "textbook" position to extreme constellations (i.e. the hyoid bone can move in a quite anterior-inferior position to the front of the thyroid cartilage).

The specific articulatory features resulting in changes of the vowel's formant values, and the raised f_0 associated with wider harmonic spacing (which means limited resolution on conveying the transfer function of the vocal tract) have the effect of reducing the acoustic vowel space with ascending pitch, and producing acoustically similar vowel qualities at the higher vocal range (Scotto di Carlo and Germain, 1985; Dowd et al., 1998; Joliveau et al., 2004; Millhouse and Clermont, 2007; Wolfe et al., 2009). In addition, some research also implied that there seem to be learned relationships between f_0 and the formant frequencies, which support human speech processing to distinguish vowel qualities, so the changes in these relations presumably distract the perceptual mechanisms in some extent (Assmann et al., 2002). At the higher boundary of this tuning (reaching a certain f_0 at about 800–900 Hz, although according to Watson's description (2009), it already happens at 698 Hz) singers tend to use a single canonical (wide open) vocal tract shape while producing all the vowels (Millhouse and Clermont, 2007; Bresch and Narayanan, 2010), therefore decreasing the distinction of acoustic vowel space not only in perception, but in production as well.

Three earlier studies regarding the perception of the high-pitched sung vowels described the increasing number of errors in identification with raising the fundamental frequency of the singing voice. Gottfried and Chew (1986) revealed that back vowels are more often misidentified than front vowels, and the different phonatory modes (so called "registers") have an important effect on the perception of the vowel sounds as well. According to Scotto di Carlo and Germain (1985), the vowels not properly identified are mostly rounded and closed, and are generally confused with open and central ones, in particular [a]. Hollien et al. (2000) described these tendencies of confusion as shifts towards vowels with higher F_1 (which

practically means a more opened configuration on the articulatory domain). Besides the latter two papers, no data on tendencies of errors occurring in misidentification were presented. In any other investigations, there are just assumptions posed (based on purely acoustic and articulatory data), implying that due to more open articulation at high pitch, vowels appear as more open in perception, too. However, it is well-known from the literature that production and perception have a non-trivial relationship, and it is a matter of agreement that any supposition of this kind has to be verified perceptually.

Nevertheless, the acoustic and percetputal tendencies of the changes of vowel production presented above are highly language dependent, since the vowel inventory differs among languages. Thus, the purpose of this study is to examine the acoustic and articulatory features of the production of the singing voice for the first time in Hungarian. The main questions are the following: 1) Is it possible to distinguish different vowel qualities in Hungarian at a relatively high pitch? 2) Which are the critical values of f_0 in the perception of sounds produced in the higher ranges of the singing voice? 3) What are the tendencies for indentification errors for high-pitched sung vowels and what articluatory background can be hypothesized for these confusions? 4) Which are the critical values of f_0 in the formant tuning with increasing the fundamental? 5) What happens to the vowels' other formants during singing?

The acoustics of children's sung vowels are not well known yet for any language, but for Hungarian, even the acoustics of children's speaking voice is under-researched (see e.g. Gósy, 1984; Deme, 2012b). However, the short vocal folds and the generally smaller vocal tract (therefore higher values of resonances and formants of speaking voice) of children imply the supposition that sung vowels and acoustic vowel space behave differently (from adults) in their singing production (e.g. because of high F_1, no tuning is necessary in case of /a:/). Moreover, it is also not clear what differences the lack of many years of training can create with regard to the energy of the spectral components of the child's voice (i.e. the vocal efficiency or loudness). To examine these questions, a pilot study was carried out on an 8-year-old girl's sung vowels. In the following sections the results of these studies and further questions are presented.

2 Pilot studies

In the ongoing work, three pilot studies have been carried out this far. In Experiment 1, the formant frequency changes and their effect on speech processing were analysed. Therefore acoustic analysis and perception tests were performed on 9 Hungarian vowels (which can be uttered with an extended duration without changing the vowel quality itself). Since the results were not entirely in agreement with previous findings described for other languages, the effect of consonantal context was tested in Experiment 2. Experiment 3 was a pilot study regarding formant tuning and vowel space reduction in a trained child's singing productions.

2.1 Subjects, material and method

As it was demonstrated earlier, vowel identification is better in consonantal context (Strange and Verbrugge, 1976). On the other hand (with regard to singing), high pitch can be achieved the most easily (whitout "forcing") while pronouncing vowels in vowel-like (e.g. nasal) context (Kerényi, 1959). Therefore for Experiment 1, nonsense *m*V*n* utterances were recorded, where the vowels /ɔ, aː, ɛ, eː, iː, oː, øː, uː, yː/ in a nasal context were sung by a professional soprano singer (age 50) at a comfortable loudness, at the steady-state f_0 values of 500, 550 and 650 Hz. For the acoustic analysis, the singer's spoken vowels (f_0 ~ 200 Hz) were used as a reference. During the perception test, the subjects' (4 males, 6 females) task was to listen to the presented sequences, and fill in the blanks left for the vowel between the given consonants on an answer sheet. Vowels in other consonantal contexts were also recorded and presented as distractor stimuli. For analysis, 36 sequences were used (1 context × 9 vowels × 4 fundamental frequencies =).

Since in the first study, disagreements were found with the earlier demonstrated confusion tendencies, the second experiment did not include just nasals, but voiced and unvoiced fricatives /ʃ, ʒ/ were also recorded (in *m*V*n*, *s*V*s*, *zs*V*zs* sequences) at the fundamental values of 500, 550, 600, 650 Hz and in speech produced by the singer from Experiment 1. The listeners (10 females, 5 males) (after hearing the whole stimulus) were asked to click on the vowel's orthographical symbol displayed on the computer screen. For analysis, 270 sequences were used (3 context × 9 vowels × 5 fundamental frequencies × 2 repetitions =). In both cases, the listeners were non-trained subjects, since we wanted to demonstrate the case of an average member of the opera audience.

In the third study, an 8-year-old girl's sung and spoken /aː, iː, uː/ vowels (in *l*V context) and three folk songs were recorded. The participating child was attending music school, and she was at the beginning of her training. The *l*V sequences were uttered in an ascending and descending scale from F3 (175 Hz) to F5 (698 Hz)[1]. For the acoustic analysis, we recorded a 30-year-old soprano's vowels for comparison. It is an important thing to emphasize, that this time, equally tempered musical notes were used (and not just a scale with phisically equidistant frequencies with no regard to musical conventions or hearing).

All of the recordings were carried out in a soundtreated room, and digitized at 44.1 kHz. The formants were determined by Fourier analysis (FFT) using Praat (Boersma and Weenink, 2011) and Wavesurfer (Sjölander and Beskow, 2009) at the middle of the vowel duration. Considering the difficulties of estimating formant frequencies from the output signal at high pitch (see 1st sec. 4th par.), it has to be

[1] In this paper musical notes are reffered to by their conventional musical names according to the Acoustical Society of America (Young, 1939). To clearly distinguish between formant frequencies and pitch values, we refer to formants by numbers in index (e.g. F_2) and to musical notes with numbers of normal size (e.g. F2).

emphasized that the presented formant values measured in these cases can only refer to the frequency of the enhanced harmonics (that possibly lie in the bandwidth of the corresponding formant), and may not be the center of the formant in question. (Similar restrictions are required when interpreting the results of Hollien et al., 2000) The listening tests were presented under headphones.

2.2 Results

2.2.1 Experiment 1

In Experiment 1 (Deme, 2012a), the results of the perception test showed a non-monotonic yet descending trend for correct identification percentages with ascending f_0 (Fig. 1). This means there is a reduction of intelligibility proportional to pitch. The cause of the sudden peak at 550 Hz is not clear yet, but it might be the effect of the so-called "register transition". The registers divide the tonal scale into pitch intervals, which are produced with the same phonatory mode, but at register transitions,changes in phonation can be observed (Titze, 2008). It was revealed that while the identification rate decays at the highest portion of a register, reaching the next (upper) one (with more optimal phonatory position) causes improvement in maintaining vowel intelligibility (Scotto di Carlo and Germain, 1985). Thus, a drop and jump in the identification rates can be the mark of a switch in production mode while raising pitch. Since there is no acceptable agreement on its acoustic properties, the effect of transition cannot be analyzed directly from objective acoustic data. One way of assessing its presence can be perceptual assessment carried out with trained listeners (singers, singing teachers), which is planned to be done in the future.

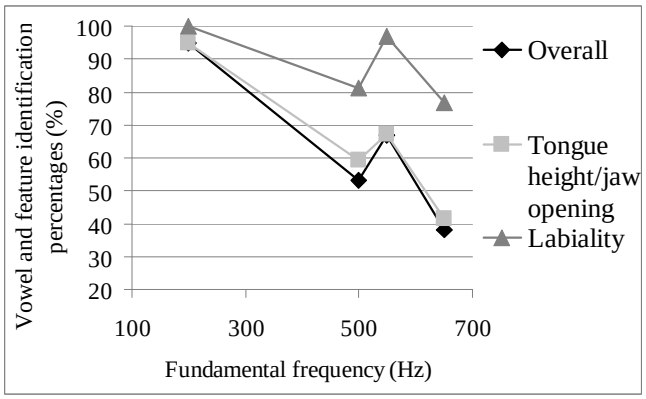


*Figure 1.*Vowel and feature[2] identification percentages in the function of fundamental frequency.

[2] Examples: along the [close] dimension /uː/ is close, /oː/ is close-mid, /ɔ/ is open-mid, /aː/ is open; along the [labial] dimension /oː/ and /uː/ are labial, /eː/ and /iː/ are non-labial.

As it it seen in Fig.1 labiality is more resistant to pitch than tongue height ($F(2) = 8.34$; $p = 0.02$), which is inconsistent with data for French (Scotto di Carlo and Germain, 1985). It also can be seen that the percentages of correct perception of the feature [close] are much higher when the vowels are spoken or produced at the lower f_0 values of singing (500 and 550 Hz). This finding is not surprising, as numerous studies have already shown that jaw opening is inversely proportional to f_0 (e.g. Sundberg, 1969, 1987; Austin, 2005; Bresch and Narayanan, 2010). That is, the higher the f_0 is, the lower the tounge/jaw is positioned. However, the types of confusions, the vowel qualities involved, and the percentages of correct recognition in detail practically show the opposite of the two already available descriptions as demonstrated in Table 1 and Fig 2.

*Table 1.*Percentages of correct identification per vowel per fundamental frequency.

f_0	Recognition percentages								
	aː	ɔ	oː	uː	øː	yː	ɛ	eː	iː
Speech (~200 Hz)	100%	90%	100%	95%	100%	100%	100%	100%	71%
500 Hz	22%	77%	62%	85%	4%	50%	73%	24%	90%
550 Hz	100%	71%	55%	43%	50%	71%	95%	60%	58%
650 Hz	26%	9%	30%	43%	0%	65%	67%	38%	67%

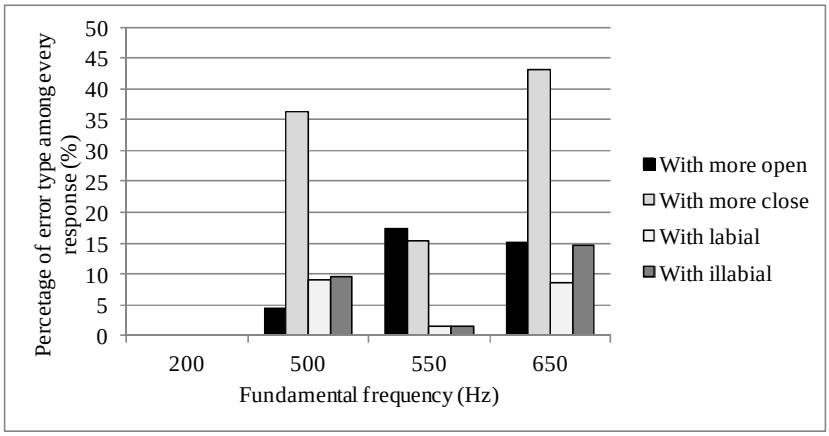


Figure 2. Confusions in vowel identification per fundamental frequency. Every response for each stimulus was worth two judgements (for the two features: [labial], [close]).

Contrary to what was demonstrated by Scotto di Carlo and Germain (1985) and Hollien et al. (2000), here at high pitch, the highest recognition percentages were measured at close (/yː, iː/) and open-mid (/ɛ/) vowels, and the vowel with the widest jaw opening (/aː/) could only maintain its intelligibility for a smaller extent (with a high recognition rate at 550 Hz). At the highest fundamental, the recognition

percentages of the feature [close] were 58.5% for close, 22.7% for close-mid, 37.2% for open-mid vowels and 26.1% for the open /aː/.

To analyze the subjects' responses, confusion matrices were constructed: one matrix per each fundamental frequency (for example see Table 2.). As the matrices show not only the number of confusions, but the types of errors that occur as well, this can be an efficient way of summing up the results.

Table 2. Example of a confusion matrix at the highest sung fundamental frequency (f_0 = 650 Hz). Number of occurences of the stimulus/response pair is indicated in the corresponding box of the matrix. For example: the vowel [aː] was mistaken for [ɔ] 12 times, for [oː] twice, and for [uː] 3 times at 650 Hz.

		Response								
		aː	ɔ	oː	uː	øː	yː	ɛ	eː	iː
Stimulus	aː	6	12	2	3					
	ɔ		2	7	13					
	oː		3	6	11					
	uː		9	3	9					
	øː					0	1	2	4	15
	yː						15		5	3
	ɛ							14	3	4
	eː							1	9	14
	iː								7	14

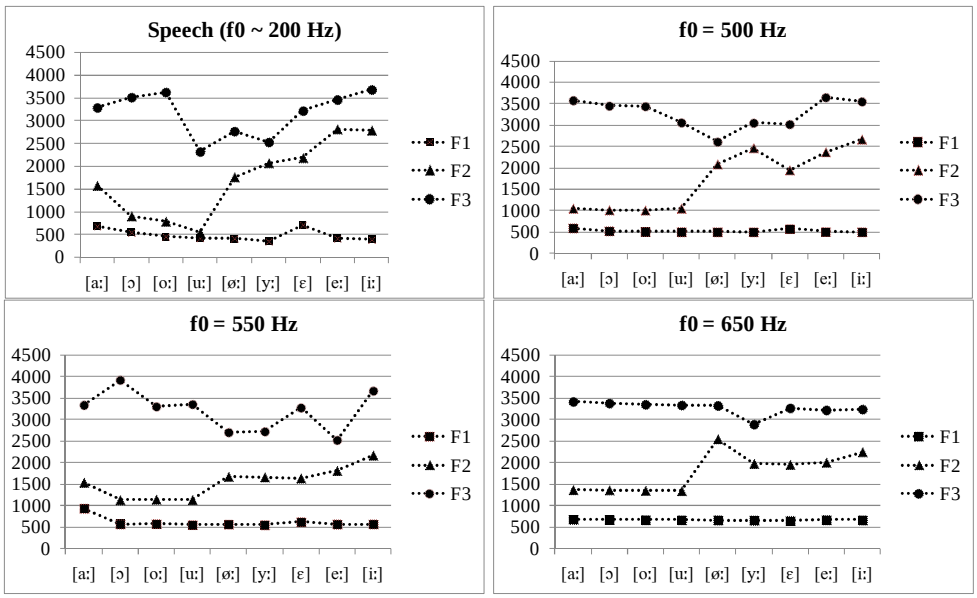


Figure 3. Measured first three formant frequencies of each vowel per fundamental frequency.

Despite the assumption, that the singer tends to articulate close and labial vowels as more open and non-labial with increasing pitch, the greatest proportion of error types in perception seems to be somewhat the opposite: not properly perceived vowels tended to be indentified as more closed sounds, in particular as /iː/ (27% out of the total number of mistaken vowels) (Fig. 2). The second most frequent vowel in the hierarchy of mistakes was /ɔ/ (16%), which might be a language or speaker specific articulatory feature of singing but still meets the expectation for vowel production with opened jaw. But as the most close vowel, /iː/ cannot fit the notion of earlier demonstrated articulation tendencies in any way.

In agreement with previous studies, the acoustic (i.e. frequency structure) differences of distinct vowel qualities are reduced with ascending pitch (Fig. 3). As the f_0 reaches the average value of F_1, the F_1 can not be distinguished from the raised pitch any more; this becomes common through the whole vowel spectrum[3]. (For certain cases (e.g. in the case of /ɛ/) the tuning seems to begin before the f_0 would exceed the vowel's F_1.) Since, the first spectral maxima and the f_0 coincide on the vowel's spectrum at any sung pitch, it can be assumed, that the singer tends to shift F_1 to match f_0 as expected. This tuning implies increase of the jaw opening with f_0 on the articulatory domain. However, the effect of more open articulation was not found in the results of the perception test.

It seems that at the higher sung f_0s, F_2 still remains as a cue for vowel frontness, that is, the back–front distinction seems to be the most resistant of all the articulatory and acoustic changes the vowels undergo. Not finding any example of back–front confusions in identifying the vowels at the examined f_0 scale confirms this observation. (Note that according to the quantal theory, the Sg_2 occuring at about 1400-1600 Hz for adult speakers is a natural separator for this feature. [see Stevens 1989 and Section 3 of the recent paper.)

As hypothesized, reduction and a categorical shift towards the vowel quality of [aː] can be observed on the acoustic vowel space (consisting of the most spaced vowels /aː, iː, uː/ in $F_1 \times F_2$ domain) (Fig. 4).

Considering the remarkable decay in acoustic vowel-differentiation with ascending pitch, the low recognition percentage of overall vowel identification (38%) at the highest pitch is not surprising. However, the appearance of the most dominant type of error (mistakes for /iː/, 27%) are not sufficiently supported by the acoustics. (As for the back vowels the high percentage of confusions with /ɔ/ [16%] practically meet the assumptions of mistakes for /aː/.)

Since the acoustic data coincide with earlier description, but in the perception tests some disagreements were found (precisely the types of errors and vowel qualities involved), in Experiment 2 the contextual effect on the intelligibility of

[3] However, considering the known limitations of the FFT analysis, it can not be excluded that the tuning might make F_1 to appear slightly higher than the f_0, as some of the authors suggest (i.e. Titze and Worley, 2009).

sung vowels was tested (Deme, 2011a). It was also a question, whether the nasal context of a sung vowel decreases the intelligibility as suggested before (Rosner and Pickering, 1994).

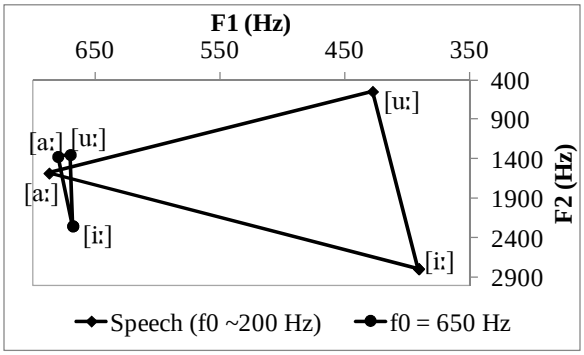


Figure 4. The acoustic vowel space consisting of /aː, iː, uː/ in speech and in singing at the highest fundamental frequency.

2.2.2 Experiment 2

In Experiment 2, the recognition percentages decreased monotonically (fell beneath 50% at 550 Hz) and appeared to be the highest for /aː/ on each fundamental frequency. The overall rate of correct identification of this vowel was 78%. This finding fits the assumption, that the opening of the jaw is enhanced while singing, therefore the intelligibility of open and mid-open vowels are the easiest to maintain even at higher pitch. At the same time, the most frequent confusion occurring during misidentification of vowels was neither with the presumed /aː/, nor /ɔ/ or /iː/ as observed in Experiment 1, but with /yː/ (24% of all mistakes) followed by /øː/ (18%) (in particular in the case of /øː, iː, eː, ɛ/).

In voiced and unvoiced fricative contexts practically the same tendency appeared as in nasal contexts: the singer mostly tended to articulate vowels which were perceived as having a more closed jaw instead of increasing the opening (Fig. 5, 6). Labiality-related confusions were less frequent than those concerning jaw opening.

Confusions with more closed sounds were dominant in every context, and at every fundamental frequency investigated. Thus, no differences were found in the function of pitch or context in this domain.

To inspect the contextual effect on vowel intelligibility, we totaled the number of confusions per consonantal environment (Fig 6), and ran the χ^2 test to evaluate the degree of differences. The statistical analysis showed significant deflection among the three types of carrier sequences ($\chi^2 = 8.511$, $df = 2$, $p = 0.014$), but as it is seen in Fig. 6, the highest identification rate characterized not the fricative, but the nasal context. The result can be interpreted as it is not only easier to sing vowels between nasal consonants at a relatively high pitch, but also efficient with regard to the matter of maintaining distinct vowel quality.

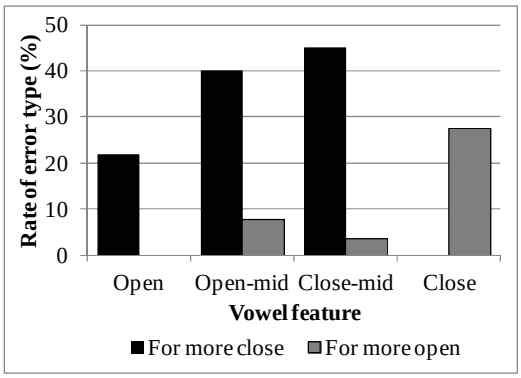


Figure 5. Percentages of error types concerning jaw opening in the function of the vowel feature [close].

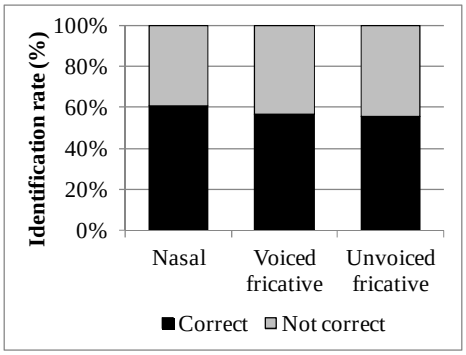


Figure 6. Correct vowel identification in terms of consonantal environment types

2.2.1 Experiment 3

In Experiment 3 (Deme, 2011b), an 8-year-old girl's sung vowels were studied. The vowel space measured in speech was greater (together with higher F_1 and F_2) than in the case of the adult woman (Fig. 7).

The sung material revealed that no vowel space reduction occurred at higher fundamentals in the child's singing (as expected from adult's data), and even a slight increase in vowel spacing was present. This change may be the result of more open articulation in the case of /aː/ and more fronting in the case of /aː, iː/. Since it was not possible any more to distinguish F_1 from f_0 for /iː/ and /uː/ at A4 (440 Hz), it can be assumed that F_1:f_0 tuning seems to begin at this fundamental frequency, but no direct tuning appeared while producing /aː/, since its average F_1 lies far above the f_0 values used by the child in singing. Therefore, the distinction among vowels seems to be preserved in the acoustic domain at higher registers as well, but (so far) there is no evidence for this separation from a perceptual point of view.

Preliminary results suggested that children's sung vowels contain less energy in the higher frequencies, so the power of the voice might be more limited in loudness than in adults' singing activity. Since this would obviously have serious

concequences on training and learning methods, it needs proper validation in the future.

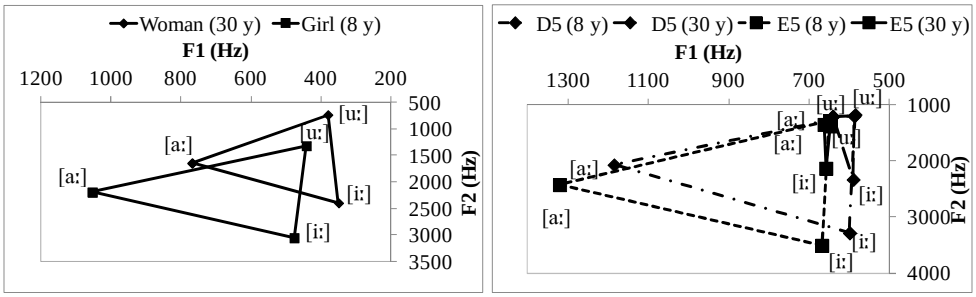


Figure 7. The acoustic vowel space of an 8 year old girl compared to a 30 year old soprano in speech (left), and at D5 (= 588 Hz) and E5 (= 659 Hz) sung fundamentals (right)

3 Discussion and future work

The aim of the author's ongoing PhD project is to determine the effect of the articulation of singing on Hungarian vowels in both the acoustic and the perceptual domains (and – to a certain extent – the effect of age). The three experiments carried out so far answered the previously formulated questions in the following way.

1) In Experiment 1 the reduction of vowel space and assimilation of vowel qualities was found to be in agreement with previous studies for other languages. However, total reduction of percpetual vowel differentitation was not observable in our material.

2) The rate of correct recognition decreased gradually with ascending pitch. The recognition percentage for the vowels produced at a relatively high pitch (650 Hz) was 38% for Experiment 1 and 37% for Experiment 2.

3) There was a tendency for the misidentification of vowels with more closed sounds (in particular as /iː/ in Experiment 1 and /yː/ in Experiment 2). This result is partly inconsistent with earlier implied tendencies, and could suggest a more closed articulation in the higher registers (in particular, in the case of front vowels). In Experiment 1 at the highest pitch, it was not open, but close sounds (/yː, iː/) that maintained their intelligibility the most. In Experiment 2, the influence of nasal context on the error types occurring in sung vowels' perception was hypothesized, but it was not confirmed. (Moreover, even some positive effects of nasal environments on vowel intelligibility were demonstrated. This finding supports the empirically established habit of singers for often using nasals for vocal exercising and training.) In this study, the error types and sustainability of vowel features seemed to be much more in line with previous research. Though the seemingly unexpected results need further parsing, possible explanations can be formulated already. Some of the inconsistencies found can be the consequence of the Hungarian vowel system. As an example, it is easy to see that /ɔ/ having three minimal pairs

differing in just one articulatory feature can dominate the hierarchy of errors much more than the expected /a:/ (in Experiment 2), which has none (as it differs from other vowels in two features or even more stages of "closedness"). In a language that has more front vowels than back ones, it is also obvious that this ratio will be represented by the higher number of errors for front vowels as well (at least at relatively lower pitches where there is no total vowel space reduction yet). The high number of mistakes for more closed sounds might be accounted for by the influence of child voice. Since the f_0 of children's speech and the corresponding formant values are normally higher than in adults' vowels, it is possible that the practice an average listener has in perceiving high-pitched sounds makes the processing system expect the formant values to be high as well. However, despite raising some of the formants observable in singing, adults' sung vowels can never have as high frequencies as it would be expected in child speech. Thus the high f_0 accompanied by relatively lower formants can cause the impression of more closed and more back articulation – while the objective acoustic data show no sign for these tendencies. This is possibly the case in the high number of mistakes for the closed vowel /y:/ in Experiment 2 or in the high recognition percentages for /y:/ and /i:/ in Experiment 1 and for /u:, y:/ in Experiment 2. Last but not least, the limitations of the results quoted from the reference literature should be noted. First, the study by Scotto di Carlo and Germain (1985) used only one singer and the same amount of material as the present one, which obviously means that the inconsistencies of their results and those presented here lie not only in the setting of this investigation, but might likely be the result of their design. Second, the paper written by Hollien et al. (2000) did not publish any exact data on the tendencies in question, just reported on some agreement with previous findings of Scotto di Carlo and Germain. Since no percentages or numbers are given in detail, no exact comparison of results is possible, and the report can only be handled with reservations. As for the childrens' sung vowels, according to our results, no vowel space reduction occurred in the acoustic domain. Although the results suggest greater vowel distinction for high-pitched singing, it still needs perceptual validation along with the issue of vocal power, which seems to be limited for children and thus can have serious consequences for the methods of training.

4) The formant tuning in adult singing began from the lowest sung fundamental frequency studied (500 Hz), and caused acoustically similar qualities within the groups of back and front vowels, but complete vowel space reduction was not reached by f_0 = 650 Hz, therefore no back–front mistakes were found. As for the child, no direct F_1:f_0 tuning was found for /a:/, but it was observable in other vowels (with lower F_1) at the pitch value of G4 (392 Hz) and above. Along with less vocal efficiency, the features of tuning (critical f_0 and tendencies) seem to be two of the most important differences between the adult and child performers.

5) As for the higher formants, the predicted formant compression and assimilation within front and back vowel groups was observable.

In addition to the above-mentioned topics, several further questions arose during the investigation. One of these is the problem posed by the appearance of the subglottal resonances in the production of sung vowels. According to Stevens' quantal theory (1989), the subglottal system, having its own resonances (Sg_1, Sg_2…) influences the speech signal, which has (among others) an important role in supporting the distinctive features of speech sounds. As an example, Sg_2 is believed to be a cause of natural separation between back and front vowels, as Sg_1 is between high and low vowels. Since normally we have no direct muscle control over our subglottal system, the values of Sg frequencies are roughly constant. However, as we already know, in singing the height and shape of the larynx (and the tension of the vocal folds) can change considerably, therefore the values and role of Sg seems to be problematic. Do Sg_1 and Sg_2 maintain their natural frequencies or function in vowel differentiation at higher f_0 as well? As it was mentioned in Experiment 1, in view of the formant structure of back and front vowel groups, we suppose so. Our preliminary results (Gráczi and Deme, 2011) show that although measuring the Sg values is almost as difficult as measuring the formants at higher registers, it seems that Sg_1 and Sg_2 slightly shift upwards with ascending f_0 (probably as a result of adjusting the larynx height or changing the phonatory position of the vocal folds), therefore its distinctive function supposedly can remain (to a certain degree) in the higher registers as well.

As the studies show, vocal tract adjustments observed in singing often make vowels ambiguous, but relatively invariant information provided by consonant articulation still can make the sung text intelligible. Therefore accurate consonant articulation in training of singing seems to be a significant factor to enhance.

References

Austin, S.F. 2005. Jaw opening in novice and experienced classically trained singers. *Journal of Voice* 21(1), 72-79.

Assmann, P. F., Nearey, T. M. and Scott, J. M. 2002. Modeling the perception of frequency-shifted vowels. *Proceedings of the 7th International Conference of Spoken Language Processing*. 425-428.

Boersma, P. and Weenink, D. 2012. Praat [Computer program] (Version 5.3). http://www.praat.org, accessed on 25 May 2012.

Bresch, E. and Narayanan, E. 2010. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *Journal of the Acoustical Society of America* 128(5), 335-341.

Deme, A. 2011a. Az énekelt magánhangzók észlelése réshangkörnyezetben. [Perception of sung vowels uttered in fricative context.] In Váradi, T. (szerk.) *V. Alkalmazott Nyelvészeti Doktoranduszkonferencia*, Budapest: MTA Nyelvtudományi Intézet. 16-28.

Deme, A. 2011b. Egy nyolcéves gyermek énekelt és beszélt magánhangzóinak akusztikai jellemzői. Esettanulmány. [Acoustic features of an 8-year-old girl's sung vowels.] *Alkalmazott Nyelvtudomány* 11(1–2), 169-188.

Deme, A. 2012a. Az énekelt magánhangzók fonetikai elemzése. [Phonetic analysis of hungarian sung vowels.] In Parapatics, A. (szerk.) *Doktoranduszok a nyelvtudomány útjain. A 6. félúton konferencia*, Budapest: ELTE Eötvös Kiadó. 33-46.

Deme, A. 2012b. Óvodások magánhangzóinak akusztikai jellemzői. [An Acoustic Analysis of Vowels in 6-7-year-old Hungarian Children's Speech] In Markó, A. (szerk.)

Beszédtudomány: Az anyanyelv-elsajátítástól a zöngekezdési időig.[Speech Science: From Language Acquisition to Voice Onset Time.] Budapest: ELTE és MTA Nyelvtudományi Intézete. 77-99.
Dowd, A., Smith, J.R. and Wolfe, J. 1998. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech 41*, 1-20.
Garnier, M., Henrich N., Smith J. and Wolfe J. 2010. Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America.* Vol. 127. No. 6, 3771-3780.
Gósy, M. 1984. Hangtani és szótani vizsgálatok egy hároméves gyermek nyelvében. [Studies on phonetical and lexicological features in 3-year-old children's speech.] *Nyelvtudományi Értekezések 102*. Budapest: Akadémiai Kiadó. 19-42.
Gósy, M. 1987. A formánsszerkezet változásának hatása a magánhangzók felismerésére. [The effect of changes in the formant structure of vowels on vowel perception.] *Magyar Nyelv* VOL LXXXII No. 2. 49-59.
Gottfried, T. L. and Chew, S. L. 1986. Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America* 79(1), 124-130.
Gráczi, T. E. and Deme, A. 2011. A szubglottális rezonanciák megjelenése az éneklésben. (talk) [Subglottal resonances in singing.] *XIII. Pszicholingvisztikai Nyári Egyetem.* Balatonalmádi, 22–26 May 2011.
Hollien, H., Mendes-Scwartz, A. P., and Nielsen, K. 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14(2), 287-298.
Hurme, P. and A. Sonninen. 1995. Vertical and saggital position of the larynx in singing. In Elenius, K. and Branderud, P. (eds) *Proceedings of the XIII International Congress of Phonetic Sciences*. 214-217.
Joliveau, E., Smith, J. and Wolf, J. 2004. Vocal tract resonances in singing: the soprano voice. *Journal of the Acousical Society of America* 116(4), 2434-2439.
Kerényi, M. Gy. 1959. *Az éneklés művészete és pedagógiája.* [The Art and Teaching of Singing.] Budapest: Zeneműkiadó.
Millhouse, T. and Clermont, F. 2007. Acoustic description of a soprano's vowels based on percpetual linear prediction. *16th International Congress of Phonetic Sciences*, Saarbrücken, 6-10 August, 2007. 901-904.
Neary, T. M. 1989. Static, dynamic, and relational properties in vowel perception, *Journal of the Acoustical Society of America 85*, 2088-2113.
Peterson, G. E. and Barney, H. L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24(2), 175-184.
Rosner, B. S. and Pickering, J. B. 1994. *Vowel Perception and Production.* Oxford: Oxford University Press.
Scotto di Carlo, N. and Germain, A. 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica*, 42(4), 188-197.
Sjölander, K. and Beskow, J. 2009. Wavesurfer [Computer program] (Version 1.8.8). http://www.speech.kth.se/wavesurfer, accessed on 25 May 2012.
Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* Vol. 17. 3-46.
Strange, W. and Verbrugge, R. R. 1976. Consonant environment specifies vowel identity. *Journal of the Acousical Society of America* 60(1), 213-224.
Sundberg, J. 1969. Articulatory differences between spoken and sung vowels in singers. *STL-QPSR* 10(1), 33-46.
Sundberg, J. 1987. *The Science of the Singing Voice*. Illinois: Northern Illinois University Press.
Titze, I. R. and Worley, A. S. 2009. Modeling source-filter interaction in belting and high-pitched operatic male singing. *Journal of the Acoustical Society of America.*126(3), 1530-1540.
Titze, I. R. 2008: Nonlinear source-filter interaction in phonation – theory. *Journal of the Acoustical Society of America* 123(5), 2733-2749.

Watson, A. H. D. 2009. *The Biology of Musical Performance and Performance Related Injury*. Lanham: Scarecrow Press.
Wolfe, J., Garnier, M., and Smith, J. 2009. Vocal tract resonances in speech, singing and playing musical instruments. *Human Frontier Science Program Journal 3*, 6-23.
Young, R. W. 1939. Terminology for logaritmic frequency units. *The Journal of the Acoustical Society of Americ*a 11(1), 134-139.

INCREASING THE NATURALNESS OF SYNTHESIZED SPEECH

Tamás Gábor Csapó
BME TMIT
e-mail: csapot@tmit.bme.hu

Abstract

In my ongoing PhD work, I am doing research in the field of text-to-speech (TTS) synthesis, particularly statistical parametric speech synthesis. For increasing the naturalness of synthesized speech, three main topics are dealt with: 1) introducing pitch variability, 2) novel excitation modeling and 3) investigating the role of subglottal resonances. 1) By investigating the fundamental frequency of speech, variability of pitch was modeled and the prosody component of a speech synthesizer was improved. Contrary to the traditional deterministic prosody models, we proposed a method to assign several pitch variants not differing in meaning for synthesized sentences, thus making the pitch component of speech synthesis more variable over longer passages. 2) In a separate experiment, the "buzzy" quality of parametric speech synthesis was reduced. Source-filter decomposition was used to obtain the speech residual signal and a novel codebook-based excitation model was proposed for use in the statistical parametric text-to-speech synthesis framework. This method uses phoneme-dependent residual frames which is an improved modeling technique compared to similar methods. 3) The role of subglottal resonances (SGRs) on speech production was investigated. We have shown that the SGRs have important phonological effects in the Hungarian language as well. Our future goal is to model the influence of the subglottal tract in speech synthesis, which is not explicitly modeled in the traditional source-filter model. Our results contribute to make synthesized speech more natural. Variable pitch has been shown to improve the naturalness of synthesized speech in a specific scenario. The novel excitation model can produce similar quality speech to other vocoders, moreover it will be possible to model different voice qualities with this method.

1 Introduction

Human speech carries large quantities of information. This can be measured objectively through physical parameters which can be matched to subjective, audible properties. For example, the physical parameter fundamental frequency (F0, frequency of vibration of the vocal folds during voiced speech) corresponds to the subjective pitch. In text-to-speech (TTS) synthesis, we can model and modify these physical properties in order to create speech that is similar to human speech.

State-of-the-art text-to-speech synthesis is based on statistical parametric methods. Particular attention is paid to Hidden Markov-model (HMM) based text-

to-speech synthesis (Zen et al., 2007). Most TTS techniques can produce good quality and highly intelligible output. Recent studies showed, however, that current speech synthesis systems are still recognized as non-human when synthesizing extended passages (Keller, 2007; Németh et al., 2007). There are a number of ways to improve naturalness of the prosody of synthesized speech: van Santen et al. (2005) minimized the prosody modification artifacts in unit selection synthesis, Díaz and Banga (2006) combined the intonation modeling and speech unit selection, while Keller (2007) analyzed perceived rhythm to make synthesized speech less robotic. Identical or very similar pitch contours of successive sentences make the synthetic speech monotonous when synthesizing longer passages of text. The first goal of our work was to design a novel prosody module capable of generating more natural pitch contours and introducing variability over successive sentences.

The source-filter model of speech separates the source (glottal excitation) from the filter (traditionally the vocal tract) (Fant, 1960). This model has been successfully applied in various parts of speech technology (e.g. in speech coding). Recent research in speech synthesis used this model in the statistical parametric framework, in the HMM-based TTS. The speech signal was decomposed to source (excitation signal) and filter, the parameters of which are modeled separately and recombined during synthesis. The second goal of our work was to improve the way in which the source-filter model, particularly the excitation signal, is used in statistical parametric speech synthesis.

It was found that the source and filter of speech are not independent and non-linear interaction between source and filter may occur (Stevens, 1998; Titze, 2008). To model the source and filter properly, it is not enough to investigate the vocal tract, because the subglottal tract (the area below the glottis) has an influence on voice as well (Lulich, 2006). This area, the lower airways (consisting of the lungs, trachea and bronchi) has resonances similar to the formants of the vocal tract. These are called subglottal resonances (SGRs). Lulich (2010) found that SGRs have phonological effects in American English. Wang et al. (2009) found that subglottal resonances can cause formant attenuation and even jumps in the second formant curve. This can be applied in a field of speech technology: they have shown that SGRs are useful in speaker normalization. Our third goal was to investigate the effects of subglottal resonances on phonological distinctive features in Hungarian, and how these could be applied in speech synthesis. In the future, particular attention will be paid to investigate the role of subglottal resonances on the excitation signal.

The following sections are organized as follows. In Sec 2) we deal with the proper modeling of pitch variability in TTS systems. In Sec 3) we improve the excitation model used in speech synthesis and Sec 4) introduces research regarding subglottal resonances in human speech. Section 5) concludes the paper and shows how the above three topics may relate to each other.

2 Modelling prosodic variability in text-to-speech synthesis

There have been only a few studies regarding variable prosody in the TTS context. Chu et al. (2006) investigated the variation of prosody in human speech using a database containing two repetitions of 1000 recorded sentences in Mandarin. A synthesis approach to variable prosody has been addressed by Díaz et al. (2006) using a unit selection TTS system. This method preserved the intonation variability of the original speaker by selecting one of several pitch candidates.

In an initial study, we have experimented with introducing prosodic variability by F0 generation in a Hungarian diphone TTS environment (Németh et al., 2007). This method generates the F0 contour for a sentence to be synthesized based on a database of natural sample sentences. However, the similarity measure used between the input text and sentences from the database was not suitable for a general speech synthesizer.

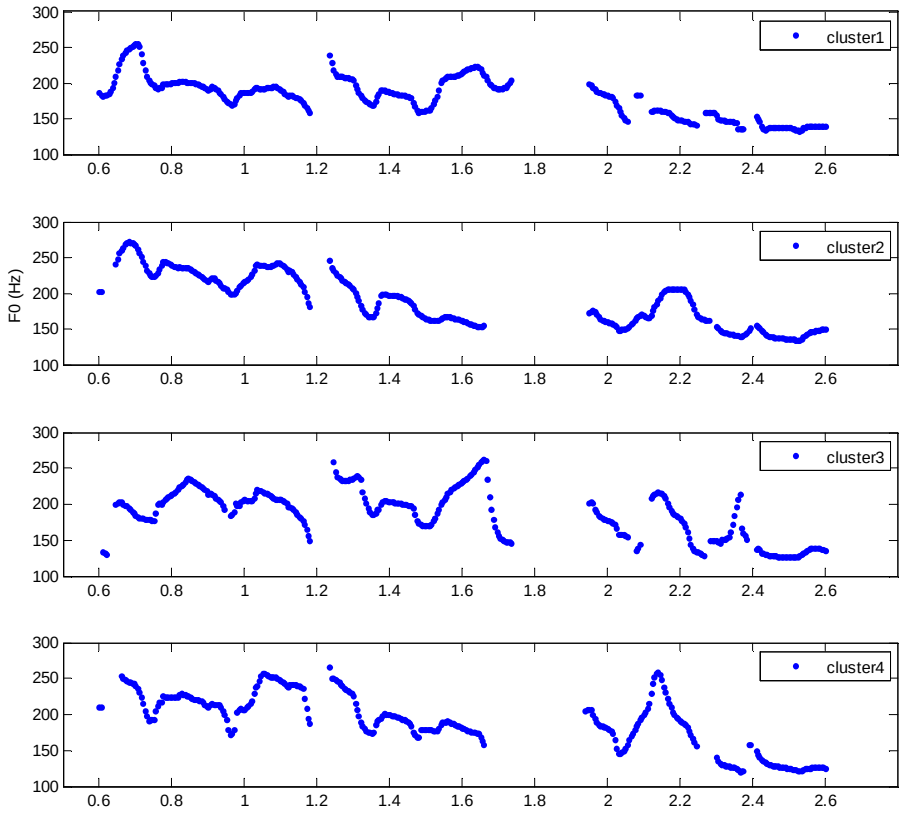


Figure 1. F0 contours of the four synthesized variants of the sentence *Zsigmond nem tagadja, hogy ő zsidó.* (’Sigismund does not deny that he is a Jew.’)

Speech synthesis research has focused recently on statistical parametric methods; particularly HMM based speech synthesis (Zen et al., 2007). Here, the basic idea is

that instead of hand-crafted rules for speech description, statistical machine learning methods are applied to analyze and synthesize speech. The parameters describing the speech signal are obtained via speech coding methods. These parameters are learned during HMM-based context clustering and assigned to the input text during synthesis. Finally, the speech is reconstructed from the parameters with a speech decoder.

HMM-based speech synthesis uses a training database of speech sentences with corresponding transcriptions. The parameters of speech (e.g., F0) are learned from this database. Typically, a few hours of speech from a single speaker is needed for acceptable quality. By splitting such a database to several subsets, and doing a separate HMM training on these subsets, the system learns different F0 models from each subset. In this way, we split a database to four subsets using the SOFM (Self-Organizing Feature Map, Kohonen et al. (1997)) unsupervised clustering method (Csapó and Németh 2011). We created four different F0 models for the HMM TTS system. During synthesis, a random F0 model is used, ensuring that the F0 contour of repeated synthesized sentences will likely be different.

Fig 1 shows an example for the results of our approach. The F0 contours of the four synthesized variants of a sentence are shown. We can see that the F0 contours are different (e.g., peaks are at different positions). Despite the differences in the F0 curve, the meaning of the four variants of the sentence is the same. Differences were audible in the pitch of the sentences according to a subjective listening test. Csapó and Németh (2011) contains an objective evaluation of this method, in which the F0 contour differences of four variants of 2000 sentences were measured.

3 Analysis and synthesis of the speech excitation signal

According to the source-filter theory, speech can be split into the source and filter (Fant, 1960). The source signal represents the glottal source that is created in the human glottis. The filter represents the vocal tract (including the mouth, tongue, lips, etc.). Traditionally, linear prediction coefficient (LPC) analysis can be used for the source-filter separation, but recently more complex and more accurate filtering methods have been used, including mel-spectrum and mel-generalized cepstrum (MGC) analysis (SPTK, 2011).

In the traditional HMM-based speech synthesis system, a very simple LPC vocoder is used for the source-filter model and an impulse sequence is used as the excitation in voiced parts, while unvoiced parts are modeled with white noise. However, this produces “buzzy” speech quality, for which HMM-based systems are often criticized. CELP (Codebook Excited LP) based methods offer the highest quality solutions to alleviate this problem (Drugman, 2011).

Fig. 2 shows the vocoding part within the HMM TTS framework. For the excitation, two types of signals are shown: 1) an impluse sequence and 2) the residual signal of speech that was obtained by MGC inverse filtering. The impulse sequence is an oversimplified model of the residual signal. The goal of our research

was to synthesize the excitation signal that resembles the properties of real residual more properly than the impulse sequence.

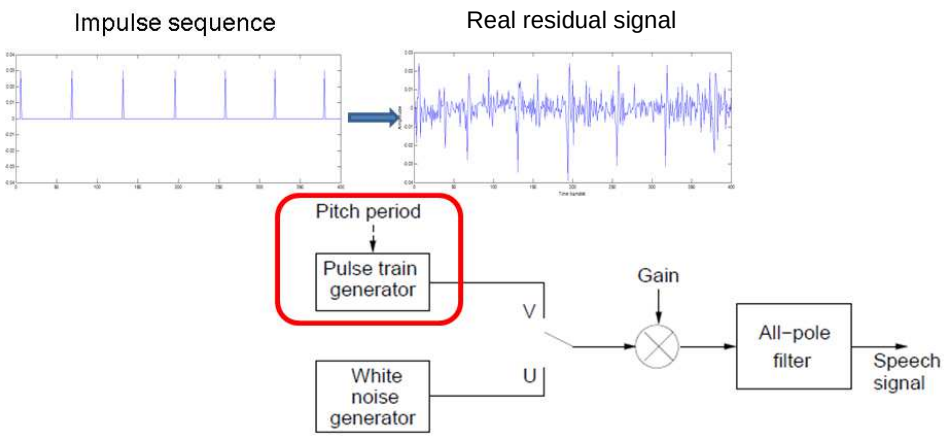


Figure 2. Vocoding within the HMM TTS framework. Two types of excitation signals are shown: the impulse sequence is an oversimplified model of the real residual signal.

Drugman (2011) was one of the the first to create such an excitation synthesis. He constructed a codebook of residual frames obtained from natural speech and used it in HMM synthesis. Cabral (2010) usesd the Liljencrants-Fant acoustic model of the glottal source derivative to construct the excitation signal. Raitio et al. (2011) used unit selection methods for the synthesis of excitation, where glottal periods obtained from real speech are concatenated resulting in a smooth excitation signal.

In our approach, we aimed to create a codebook-based excitation model that uses unit selection (Csapó and Németh, 2012). During the analysis part, the residual signal was obtained from natural speech with MGC-based inverse filtering. Starting from this signal, a codebook was built from phoneme-dependent pitch-synchronous excitation frames. Phoneme dependent frames were used, because we assumed that the inverse filtering was not perfectly decomposing source and filter, and that the residual signal may contain information regarding the phone as well. In other codebook-based vocoding methods, there is a general codebook obtained from all of the phones. Several parameters (e.g., period, T0, energy) of these frames are fed to the HMM training system. For the sentences to be synthesized, the HMM TTS assigns these parameters for each speech sound. During synthesis, phoneme-dependent excitation frames are selected from the codebook with unit selection, and concatenated to each other. After that, final synthesized speech is obtained with MGC-based filtering.

Subjective analysis of the quality that can be synthesized with this method has not been conducted yet, but our preliminary tests suggest that the quality is similar to other CELP based methods.

With this novel excitation approach, we will be able to model different voice qualities. By creating separate codebooks from breathy, whispered, or other type of speech, the method can synthesize speech with the specific voice quality.

4 Investigation of the relation between vowel formants and subglottal resonances

It has been shown for several languages that subglottal resonances play a role in dividing the frequency space of consonant and vowel acoustics into discrete regions corresponding to phonological categories. Lulich (2006) investigated American English speakers, Madsack et al. (2008) tested several speakers of two German dialects and Jung (2009) tested Korean speakers. SGRs have been reported to divide vowels into certain contrasting natural categories: low – non-low; front – back; front tense - lax. Our work aimed to consider the patterns of Hungarian vowels with regard to the SGRs in speech production and perception.

In a first experiment, we investigated the vowel space of four Hungarian speakers in nonsense word reading (Csapó et al., 2009). Subglottal resonances were measured from the accelerometer signal; the accelerometer was pressed to the neck while speaking. The results confirmed that the first subglottal resonance (Sg1) divides low and non-low vowels (in terms of the first formant, F1), while the second (Sg2) separates back and front vowels in Hungarian as well (in F2). The third subglottal resonance (Sg3) has been found to divide unrounded non-low front vowels from other front vowels. An example for this can be seen in Fig. 3. The figure shows the separating role of SGRs for a specific Hungarian speaker. The dividing line between low and non-low vowels (Sg1) is less clear compared to Sg2 and Sg3, possibly because it is more difficult to measure Sg1 than Sg2. The results are similar for other speakers as well (Csapó et al., 2009).

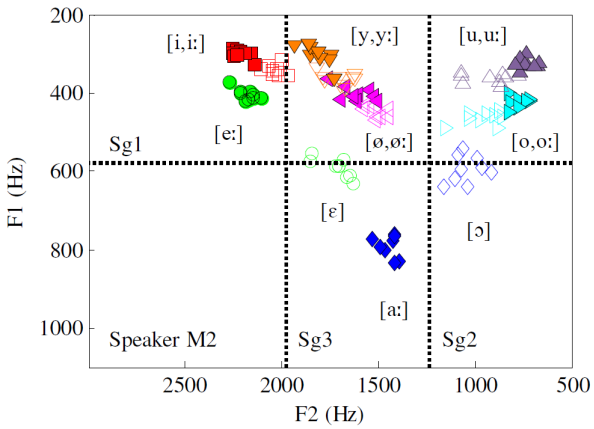


Figure 3. Formant space of a speaker with the 14 Hungarian vowels. The subglottal resonances are indicated by horizontal and vertical dashed lines.

In a second experiment, the formant spaces of six other speakers were analyzed in spontaneous speech (Csapó et al., 2011). We found that the SGR effects were less intensive in continuous speech compared to the nonsense word recordings. Particularly, data from most speakers did not support the role of Sg3.

In the next experiment, using data from the same six speakers, automatic formant-based vowel classification was applied and extended with normalization based on the subglottal resonances (Csapó et al., 2011). The input of the classification included the first two formant values of vowels extracted from spontaneous speech and the first three subglottal resonances. The target of the classification was groups corresponding to phonological distinctive features – e.g., in case of F2-Sg2 the target was 'back vs. front'. Three types of classifications were performed, of which two are shown here: 1) decision tree classification on raw formant data (without SGRs) and 2) decision tree classification on SGR-normalized formant data. Table 1 summarizes the correctly classified rates of these classifications. In 'low vs. non-low', Sg1 was used for F1 formant normalization, in 'back vs. front', Sg2 was used for F2 formant normalization and in 'front* vs. other', Sg3 was used for F2 formant normalization. According to the results of the experiment, it was shown that the knowledge of Sg2 and Sg3 may improve the accuracy of automatic vowel classification if using male and female data together in limited data contexts. Sg1 was not helpful in automatic classification of spontaneous speech.

Table 1: Result of the vowel classification experiments: correctly classified rates. Front* denotes the 'front unrounded non-low' vowel class.

	decision tree	**decision tree + SGR**
low vs. non-low	79.81%	78.09%
back vs. front	84.28%	84.73%
front* vs. other	86.95%	88.21%

Finally, in a pilot experiment, the perceived backness of the vowel [ɔ], as a function of F2, and Sg2 was investigated in a listening test (Csapó et al., 2011). C[ɔ]C transitions with different F2 frequencies at the vowel midpoint were extracted from two speakers' spontaneous recordings. Ordering the vowels by increasing F2, the results showed that for one of the two tested speakers, an abrupt increase in perceived backness of the vowel occurred when F2 was higher than Sg2. For the other speaker, a similar abrupt increase was not observed in the F2 – Sg2 relation.

It has been shown that SGRs can be applied in speaker normalization and automatic speech recognition (Wang et al., 2009; Arsikere et al., 2011). However, the usefulness of subglottal resonances in speech synthesis has been only initially investigated before. Gorbunov and Makarov (2011) modeled the subglottal region in an articulatory speech synthesizer by simulating the influence of the trachea, bronchi and and lungs. Hiroya (2011) introduced a method to remove the effect of subglottal resonances in speech signals for estimating a vocal-tract spectrum, and showed its

accuracy in Japanese speech synthesis examples. It is possible that by modeling subglottal resonances explicitly in statistical parametric speech synthesis, the quality of this could be improved.

5 Summary

Speech synthesis has been tackled recently using statistical methods. In this study we have experimented with introducing prosodic variability by F0 generation in a Hungarian TTS environment. Our method generated several F0 contour variants for a sentence, and from this a random candidate could be chosen during synthesis. According to van Santen et al. (2005), this may be an important feature of future speech synthesis.

The HMM TTS technique makes use of large speech corpora, of which the main parameters of speech are extracted and later recombined. A significant problem of state-of-the art statistical parametric speech synthesis systems is their "buzzy" quality caused by an oversimplified excitation model. As part of my research, I am developing ways to more accurately model the excitation of a Hungarian statistical speech synthesis system, thereby improving the quality of synthetic speech. Compared to other excitation models (e.g. Drugman, 2011; Cabral, 2010; Raitio et al. 2011), my model is expected to produce similar quality synthesized speech. By further improving the excitation model, we will be able to synthesize different voice qualities (e.g. breathy, whispered) as well.

Investigating subglottal resonances in Hungarian helped to understand how speech production works. We have found relationships between vowel formants and SGRs that are similar to other languages (e.g. Lulich, 2006). It is not known whether the parameters currently used in HMM TTS properly model the phenomena caused by subglottal acoustics (e.g., jumps in the F2 track and formant attenuations) in synthesized speech.

The results regarding subglottal resonances have implications for understanding phonological distinctive features, as well as applications in automatic speech technologies. The latter includes speaker normalization (e.g. Wang et al., 2009) and other related problems in automatic speech recognition. The fact that SGRs are roughly constant for a given speaker may be useful in speaker recognition as well (Arsikere et al. 2011).

In the future, I plan to investigate the effect of SGRs on the excitation signal of speech obtained by inverse filtering. Several studies (on American English and Spanish) have confirmed that there are correlations between specific properties of the speech signal and SGRs, thus an indirect estimation of the resonances can be done using microphone recordings (Arsikere et al., 2011). Testing these algorithms on both Hungarian and English recordings will help to extend the hypothesis that subglottal resonances have relevance independent of the language used. A better understanding of how SGRs affect the glottal source will suggest new ways to improve the naturalness of synthesized speech.

In my PhD work I have been doing research in several subfields of speech science. In text-to-speech synthesis, our goal is to make the synthesized speech as close as possible to human speech. With my results, future speech synthesis is expected to become more natural.

Acknowledgements

I would like to thank the editor and the two anonymuous reviewers for the valuable suggestions and comments that have improved the manuscript. This research has been partially supported by the Paelife (Grant No. AAL-08-1-2011-0001), by the CESAR (Grant No. 271022) and by the TÁMOP-4.2.1/B-09/1/KMR-2010-0002 projects.

References

Arsikere, H., Lulich, S.M., and Alwan, A. 2011. Automatic estimation of the second subglottal resonance from natural speech. *ICASSP 2011*, 4616–4619.

Cabral, J. P., 2010. *HMM-based Speech Synthesis using an Acoustic Glottal Source Model.* PhD Thesis, CSTR, University of Edinburgh, United Kingdom.

Csapó, T.G., Bárkányi, Zs., Gráczi, T.E., Bőhm, T. and Lulich, S.M. 2009. Relation of formants and subglottal resonances in Hungarian vowels. *Interspeech 2009*, Brighton, United Kingdom, 484–487.

Csapó, T.G., Gráczi, T.E., Bárkányi, Zs., Beke, A. and Lulich, S.M. 2011. Patterns of Hungarian vowel production and perception with regard to subglottal resonances. *The Phonetician* 99-100, 7–28.

Csapó, T.G. and Németh, G. 2011. Prozódiai változatosság rejtett Markov-modell alapú szövegfelolvasóval, [Prosodic Variability in Hidden Markov Model-based Text-To-Speech Synthesis]. *MSZNY*, Szeged, Hungary, 167–177.

Csapó, T.G. and Németh, G. 2012. A novel codebook-based excitation model for use in speech synthesis. *CogInfoCom 2012*, Kosice, Slovakia, accepted.

Chu, M., Zhao, Y. and Chang, E. 2006. Modeling stylized invariance and local variability of prosody in text-to-speech synthesis. *Speech Communication* 48, 716–726.

Díaz, F.C. and Banga, E.R. 2006. A method for combining intonation modelling and speech unit selection in corpus-based speech synthesis systems. *Speech Communication* 48, 941–956.

Drugman, T. 2011. *Advances in Glottal Analysis and its Applications.* PhD Thesis, University of Mons, Belgium.

Fant, G. 1960. Acoustic theory of speech production. Mouton, The Hague.

Gorbunov, K.S. and Makarov, I.S. 2011.The subglottic region in articulator synthesizers. *Journal of Communications Technology and Electronics* 56, 1504–1509.

Hiroya, S., Miki, N., and Mochida, T. 2011. Multi-closure-interval linear prediction analysis based on phase equalization. *Proc. APSIPA.*

Jung, Y. 2009. *Acoustic articulatory evidence for quantal vowel categories: The features [low] and [back]*. PhD Thesis, MIT, USA.

Keller, E. 2007. Beats for individual timing variation. In A. Esposito et al. (eds.), *The Fundamentals of Verbal and Non-Verbal Communication and the Biometrical Issue*, IOS Press.

Kohonen, T., Kaski, S., and Lappalainen, H. 1997. Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural Computation*, vol. 9, no. 6, 1321–1344.

Lulich, S.M. 2006. *The Role of Lower Airway Resonances in Defining Vowel Feature Contrasts.* PhD Thesis, MIT, USA.

Lulich, S.M., 2010. Subglottal resonances and distinctive features. *Journal of Phonetics* 38(1), 20–32.

Madsack, A., Lulich, S. M., Wokurek, W., and Dogil, G. 2008. Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs. *Proceedings of LabPhon 11*, 91–92.

Németh, G., Fék, M., and Csapó, T.G. 2007. Increasing Prosodic Variability of Text-To-Speech Synthesizers. *Interspeech 2007*, Antwerp, Belgium, 474–477.

Raitio, T., Suni, A., Yamagishi, J., Pulakka, H., Nurminen, J., Vainio, M., and Alku, P. 2011. HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering. *IEEE Transactions on Audio, Speech and Language Processing* 19(1): 153–165.

van Santen, J., Kain, A., Klabbers, E., Mishra, T. 2005. Synthesis of Prosody using Multi-level Unit Sequences, *Speech Communication*, Vol. 46(3–4), 365–375

SPTK working group. 2011. *Reference Manual for Speech Signal Processing Toolkit* Ver. 3.5, December 25, 2011.

Stevens, K.N. 1998. *Acoustic Phonetics*, MIT Press, Cambridge, MA.

Titze, I.R. 2008. Nonlinear source–filter coupling in phonation: Theory. *Journal of the Acoustical Society of America* 123, 2733–2749.

Wang, S., Lulich, S. M. and Alwan, A. 2009. Automatic detection of the second subglottal resonance and its application to speaker normalization. *Journal of the Acoustical Society of America* 126, 3268–3277.

Zen, H., Nose, T., Yamagishi, J., Sako, S. Masuko, T., Black, A.W. and Tokuda, K. 2007. *The HMM-based speech synthesis system version 2.0, ISCA SSW6*.

CONFERENCES IN 2012

During 2012, there were a number of conferences that focused on speech research. I attended three conferences of these conferences.

The first was the Workshop on Innovation and Applications in Speech Technology (IAST) in Dublin, Ireland. The goal of this conference was to discuss the results and most importantly the plans for speech technology in the future. In two days, this workshop covered a diverse array of topics: from expressive speech and multimodal applications to dialogue and human-computer spoken interaction, and of course not to forget a session that provided ideas on how to Rock the scientific style. The sound of new musical instruments was presented through synthesis, creating a new world of music, and new perspectives on the naturaleness of speech technologies. One set of presentations concerned the needs of phonetic knowledge in speech technology. The conclusion was that there is a high need for phonetic research in the investigation of articulatory processes and its modeling, leading to articulatory synthesis systems for plausibility testing and new hypothesis generation. Another group of researchers analyzed the creaky voice in speech and developed an algorithm to classify instances of this type of production in read and spontaneous speech. These authors proposed a method called Resonator-based Creaky Voice Detection (RCVD). The fundamental idea was that there is a secondary peak in the LP-residual signal during creaky voice. One of the most interesting topics was long-term speaker verification. In this presentation, the researcher tried to compensate for the effects of aging on voice.

Image from the presentation of Ingmar Steiner–Slim Ouni: Artimate: an articulatory animation framework for audiovisual speech synthesis (http://www.coli.uni-saarland.de/~steiner/pdf/IAST2012Slides.pdf)

The second conference was the 15th International Conference on Text, Speech and Dialogue (TSD) in Brno, Czech Republik. There were three primary areas presented at this conference: i) corpora, text and transcription; ii) speech analysis, recognition, synthesis; and iii) their intertwining within natural language dialogue systems. The topics covered by the presentations represented a wide range. Over five days, there were a number of diverse topics presented: form word disambiguation, key phrase extraction, emotion recognition, in-car speech recognition system, classification of healthy and pathological continuous speech, a spoken dialogue system, aggression detection, question classification, etc. One of the most interesting talks was the impact of non-speech sounds on speaker recognition. The results showed that a non-speech sound (e.g., breathing patterns) could play an important role in speaker recognition. Another interesting topic was the classification of healthy and pathological speech using a support vector model. Jitter and harmonics-to-noise measures were used. These articles are available in the conference proceedings.

Some pictures from the TSD in Brno

The 3rd IEEE International Conference on Cognitive Infocommunication (CogInfoCom) was the last conference chosen to be introduced in this short summary. The goal of this conference was to bring together several areas of science into one platform. CogInfoCom tries to establish an interaction among cognitive science, infocommunication and engineering applications. The conference's primary goal was to model the human-human and human-machine interaction. During the four days, there were many very interesting presentations and amazing demonstrations. The topics conveyed the complexity of these science areas: augmented cognition, body area network, cognitive informatics and media, cognitive linguistics, cognitive robotics, cognitive science, ethology-inspired engineering, etho-robotics, 3D visualization and interaction, human-computer and human-robot interaction, iSpace research, interactive systems engineering, media informatics, multimodal interaction, real and virtual avatars, sensory substitution and sensorimotor extension, teleoperation, virtual reality technologies and scientific visualization. Many presentations on topics of phonetic relevance were introduced at

this conference as well. For instance, Anna Esposito and her colleagues studied the visual durational cues that help native and non-native speakers to discriminate single and geminate consonants. This is a new perspective on the role of the effect of temporal factors in speech. One of the presentations showed how laughter was an important phenomenon in spontaneous speech that can be used to detect topic change. During the demonstration section, some robots and other applications could be seen at work, showing the future trends of research in this area. These articles can be found in the proceedings of the conference.

Photo from CogInfoCom

András Beke
Research Institute for Linguistics,
Hungarian Academy of Sciences,
Budapest, Hungary
e-mail: beke.andras@gmail.hu

FUNCTION OF THE SINGING VOICE

KTH-course DT 211 V of the CSC/Department of Speech, Music and Hearing Malmköping, Sweden, 28 July – 3 August 2012

The summer course *Function of the Singing Voice* has been organised by The School of Computer Science and Communication of The Royal Institute of Science since 2007. It takes place in Sandvik, Malmköping, a small, quiet and idyllic place 115 km southwest from Stockholm. The main teacher, Johan Sundberg former director of the music acoustic research group at the Department of Speech, Music and Hearing, has created a hands-on curriculum that explains how the voice functions when used as a musical instrument within the classical Western tradition and several contemporary genres. The course is divided between lectures and workshops, where participants are also given the opportunity to watch and analyze their voice and respiratory system in speech and singing.

The emphasis of this summer course was placed on how the voice works and how its timbral properties are controlled by physiological, such as breathing behavior, larynx positioning and vocal tract shaping, but room acoustics, basic sound recording technology and auditory perception of the voice were also included. Therefore, people with various backgrounds and professions (singers, speech and singing pedagogues, speech therapists, ear nose and throat doctors, and phoneticians) participated.

Lectures given on the topic of "Function", "Formants", "Tube phonation" and "Source" summed up the acoustic background, and lectures like "Functional anatomy" and "Breathing" introduced the anatomical bases needed for investigating the function of the human voice. From the second day on, current investigations and novel scientific results were presented (in lectures like "Hormones and the voice", "Voice in the choir", "Adding expressiveness to musical performance" or "Secrets of an ugly voice"). "Room acoustics" and "Microphones and microphone placement" laid the foundations for experimental use of audio recording.

The participants attended the workshops as smaller groups. They learned how to use *Respiratory inductive plethysmography* for recording breathing movement; *oscillogram* and *inverse filtering* for recording the *flow glottogram* of the pulsating glottal airflow (and also examined the differences of the functions caused by phonation types); the articulatory model *APEX* (developed by researchers of KTH) to analyze and build the vocal tract shapes of different vowels; *Madde* (also developed by KTH scientists) for synthesizing vowels with adjustable acoustic properties; *Phonetogram* for profiling the voice range (in speech and singing) and a pressure transducer with an oscilloscope and a manometer for measuring the *subglottal pressure* of singing at different levels of loudness. Finally, the

participants were also given the chance to watch the movement of the vocal folds and the larynx during phonation using a *fiberscope*.

A masterclass given by the famous opera singer Håkan Hargegård, demo lessons given by singing teacher Brian Gill and workshops held by Daniel Zangger Borch and Margareta Thalén gave the singers among the participants a unique opportunity to develop their training and teaching techniques in several genres and also balanced the interesting mixture of theory and experience defining the whole event. The course "Voice health" and "Bubble phonation" provided practical advice and useful information for those who are using or analyzing the voice as a profession as well.

The teaching staff (Christine Ericsdotter, Anders Friberg, Brian Gill, Svante Granqvist, Håkan Hargegård, Stellan Hertegård, Filipa Lã, Frank Müller, Camilla Romedahl, Gláucia Salomão, Thomas Schuback, Johan Sundberg, Sten Ternström, Margareta Thalén, Daniel Zangger Borch) consisted of musicians, singers, singing teachers, voice researchers, logopeds and ENT doctors from various universities and countries (Sweden, Germany, Portugal and the USA).

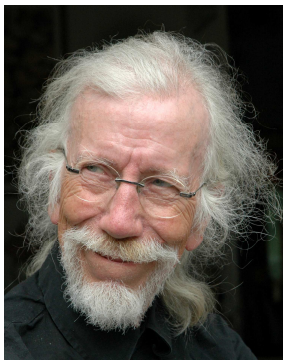
Besides the exhaustive scientific work, the summer course also included a great social program. The lessons were held in a quiet ranch in absolute isolation from 9 am to 9 pm, thus the participants from different backgrounds were not only working together in the courses, but also got included in scientific discussions and socialization during the breaks and leisure time.

The main teacher Johan Sundberg demonstrating inverse filtering at a workshop.

Andrea, Deme
Department of Phonetics, Eötvös Loránd University, and
Research Institute for Linguistics, Hungarian Academy of Sciences,
Budapest, Hungary
e-mail: deme.andrea@nytud.mta.hu

OBITUARY

In Memoriam
Johan Liljencrants (1936–2012)



Johan Liljencrants was a KTH oldtimer. His interests focused early on speech analysis and synthesis where in the 1960s, he took a leading part in the development of analysis hardware, the OVE III speech synthesizer, and the introduction of computers in the Speech Transmission Laboratory. Later work shifted toward general speech signal processing. His interests expanded to modelling the glottal system, as well as physically including glottal aerodynamics and mechanics. Johan took a leading part in the development of analysis software and hardware. Modern techniques with integrated circuits were the base for OVE III. The synthesizer was controlled by a computer program utilizing smoothed step functions in an elegant fashion. At an early stage, the importance of the glottal source was realized, and detailed work with Gunnar Fant, John Holmes, Jan Lindqvist Gauffin and Martin Rothenberg led to ways to increase the naturalness of speech synthesis considerably. This work was finally implemented as the LF-glottal source (Liljencrants–Fant glottal source), which is still a reference model and used in many research efforts and applications.

Johan was amazingly productive, not so much in terms of traditional academic reporting, but in terms of devices, methods, programs and ideas that made life so much more interesting and easy for many of us, his colleagues in the department. For four decades he was possibly the most important scientific support and discussant partner to Gunnar Fant.

After his retirement, the Fonema company web site developed into a comprehensive account of his wide interests, also outside the area of speech communication. He described himself as follows: ‘Beside his private interest in music he indulges in various handicrafts like single-line flying kite design, carpentry, sewing, bottleship building, silver forging, and gem facet cutting’ (http://www.fonema.se). His music interest not only included playing the trumpet in the department’s ‘formant orchestra’ but also constructing and building a street organ from scratch (http://www.fonema.se/organ/organ.htm).

Johan was an exceptionally talented and multifaceted person. We miss him a lot as a researcher and a fellow human being.

Rolf Carlson, Björn Granström
Royal Institute of Technology (KTH), Stockholm, Sweden
e-mail: {bjorn|rolf}@speech.kth.se

BOOK REVIEWS

Iyabode Omolara Daniel (2011): Introductory Phonetics and Phonology of English

Cambridge Scholars Publishing, Newscastle, xvi + 112 pp.
Hardback: ISBN: 9781443826389, price: £ 34.99 € 52,99

Reviewed by: **Chantal Paboudjian**
University of Provence, Aix-en-Provence, France
e-mail: ChPaboudjian@aol.com

Introductory Phonetics and Phonology of English, as stated by the author, is an "Introductory course meant to help students familiarize themselves with the basics of the English Phonetics and Phonology". Its objective is to provide a practical guide to the learner in both the theoretical and practical uses of the Phonetics and Phonology of English.

Its author, Iyabode Omolara Daniel, possesses 20 years of experience in the Linguistics of English. She is Senior Lecturer and Head of Department at the National Open University of Nigeria in Lagos, Nigeria, where she teaches courses in Sociolinguistics and Gender Studies. She has published numerous articles since 2000 mainly on Language Skills, Gender, English Phonetics and Phonology.

The 112 pages of the volume are divided into 11 short chapters ending with 3 to 5 practical questions and 3 appendices of Phonetic symbols, of the stressed and unstressed forms of some (39) grammatical words and of the International Phonetic Alphabet. The chapters can be outlined as follows:

Chapter 1. General Introduction (pp. 1-3) defines the notions of "Phonetics" and "Phonology" and explains the differences between the two terms.

Chapter 2. The Mechanisms of Speech Sounds (pp. 5-13) examines the physiological processes involved in sound production, i.e. the organs of speech, the air stream mechanism, the stages of speech production, the states of the glottis and the resonators. Six figures illustrate the chapter.

Chapter 3. Articulation of English Sounds (pp. 15-27). As indicated by its title, the chapter is devoted to the place and manner of articulation of English consonants and to the features of the vowels and diphthongs. It stresses that the nature of the major sound features of English (vowel or consonant) can be determined by Gimson and Ramsaran's questions (1989). Figures of the cardinal vowels, the pure vowels (or monophthongs) and the diphthongs of English are provided.

Chapter 4. Examples of Consonants and Vowels of English in Words (pp. 29-44) contains lists of the sounds of English, i.e. 24 consonants, 20 vowels, 8 diphthongs and 2 triphthongs, as well as details related to the different spellings of

each sound. They are followed by comments on the occurrence of triphthongs based on Arnold and Gimson's (1973) classification. The chapter closes with suggestions aimed at improving personal pronunciation practice.

Chapter 5. The Suprasegmentals (pp. 45-50) begins with a definition of the term itself followed by a discussion on the nature and form of the English syllable based on Chomsky's transformational generative grammar (1966) and on Roach (1993). The rest of the chapter deals with the determination of syllable boundaries and with consonant clusters and their possible forms according to position.

Chapter 6. The English Stress (pp. 50-64) contains an introductory section that defines the notions of stress and of stress-timed rhythm. Several pages are devoted to word stress placement according to the number of syllables and to suffixes and affixes. A section is dedicated to compound words and to the binary (noun-adjective/verb) opposition of stress. The last three pages concern sentence stress with details on the roles of lexical and grammatical words and of emphatic stress.

Chapter 7. Rhythm (pp. 65-68). The three pages of this chapter deal with rhythmic isochrony and acknowledge that the issue remains a controversial one in linguistics.

Chapter 8. In the chapter entitled **Intonation** (pp. 69-76), the term itself is introduced along with its main functions and elements, such as the intonation phrase, the boundary, and the nuclear tone. The five basic tunes of English are then presented with their functions. Emphatic contrast at the sentence level and the relationship between punctuation and intonation are then discussed. Finally, the author underlines the importance for students to master intonation due to its important syntactic and semantic functions.

Chapter 9. Minimal Pairs (pp. 77-83). This is the first of the three last chapters of the volume dedicated to the description of the phonological features of English. Daniel argues that the description of phonological features makes the identification and description of the sounds of English more accessible and therefore more interesting to language students. The chapter explains the importance of phonological features in establishing the phonemic status of sounds and other speech elements, such as the phone, the phoneme and the allophone. It includes five pages of examples of minimal pairs of English. The chapter ends with a brief mention of stress features, i.e., stress placement in some orthographically identical words is phonemic.

Chapter 10. The Phonological Features of English Sounds I (pp. 85-91) treats allophonic and allomorphic variations. The author defines these notions with series of examples and then lists the allophonic variants of some phonemes, i.e. [t], [l] and [n], and the allomorphic variants of important morphemes, i.e., the plural morpheme, the past tense morpheme, the genitive morpheme, the third person singular verb morpheme and the indefinite (a/an) articles.

Chapter 11. The Phonological Features of English Sounds II (pp. 93-104) begins with a definition of distinctive features then lists some of Chomsky and Halle's (1968) phonological features describing sounds and the major class features

(sonorant/non sonorant; vocalic/non vocalic; consonantal/non consonantal, cavity features, manner of articulation features and source features).

What strikes the reader first with this book is its clear presentation and easy to read font. As far as content, clear definitions make difficult abstract notions simple and understandable. The chapters are rather short contributing making it easy to grasp important definitions and contains useful suggestions on how to overcome learners' fear of English sounds. It is worth mentioning that an attempt is made to give detailed information on the workings of the prosodic features of English which, according to the author, constitute the most confounding aspect of the English language to students. An additional chapter on phonotactics rules would have probably been helpful.

The volume is an interesting and worthwhile resource for undergraduates but also for postgraduates who want to understand and master the phonetics and phonology of English. It will also be helpful to teachers of Linguistics who can use the basic theoretical notions in Linguistics, the suprasegmental examples, and the practice questions found at the end of each chapter. Finally English teachers as well may get more familiar with some notions that are oftentimes missing in textbooks but nevertheless important in the teaching of pronunciation.

The price may constitute an obstacle - especially for students - to actually purchasing the book. However, for departments and individuals who need a tool to quickly reinforce notions in Linguistics, it is worth considering the expense. For the sceptics or those who would like to learn a bit more about the book, a free download of the fist chapter is available online (PDF format) at the following link: http://www.c-s-p.org/flyers/978-1-4438-2638-9-sample.pdf.

References

Arnold, G. F. and Gimson, A. C. 1973. *English Pronunciation Practice*, London: University of London Press.

Chomsky, N. (1966). *Topics in the Theory of Generative Grammar*, The Hague: Mouton.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*, New York: Harper and Row.

Gimson, A. C. and Ramsaran, S. 1989. *An Introduction to the Pronunciation of English*, Kent: Edward Arnold.

Roach, P. 1993. *English Phonetics and Phonology, a Practical Course*. 2nd ed., Cambridge: Cambridge University Press.

Mohamed Embarki and Christelle Dodane (eds.) (2011):
La Coarticulation: Des Indices à la Représentation
(Coarticulation: From Signs to Representation)[1]
Peter Lang GmbH, 329 pp.
Paperback ISBN 978-3-631-57746-2, $81,95

Reviewed by: **Judith Rosenhouse**
SWANTECH Ltd. Haifa, 32684 Israel
e-mail: swantech@013.net

This book assembles 15 articles about various aspects of coarticulation in nine languages by single or several authors. The studied languages include three varieties of French and of Catalan, two varieties of Venezuelan Spanish and of Arabic, as well as English, Portuguese, Russian and Korean. Following an introductory chapter by the editors,[2] the book has five sections: 1. Theoretical models; 2. Observation and instrumentation methods; 3. Acoustic signs; 4. Perception; 5. Phonological representation and linguistic constraints. An important feature of this book is that it is the first book dedicated to coarticulation that appears in French, though about half the papers were translated from English by several of the participating scholars.[3]

In the first chapter by Mohamed Embarki and Christelle Dodane, "Coarticulation, past and present" (pp. 7-16), the editors of this volume, write that the current studies of coarticualtion differ greatly from past ones. These authors define the term coarticulation based on past literature (e.g., Kuehnert and Nolan, 1999; Hammarberg 1976; Bladon and al-Bamerni, 1976; Hardcastle and Hewlet 1999) as follows: "coarticulation reflects essentially the lack of correspondence between the basic linguistic elements, phonemes or features, and their realization" (p. 8). But they stress that the concept reflects grading of the linguistic elements, rather than just their leveling or contamination. Thus, coarticulation "has one foot in production and the other in cognition." Three points direct the structure of this volume, they write: 1. accepting the term coarticulation as designating changes of segments of a phonetic sequence dictated by the lack of correspondence between the linguistic elements and the gestures involved in their realization. 2. This lack of correspondence makes coarticulation a simple mechanical process on the realization domain. 3. Modifications are supported by a search for balance between the

1 The book is written in French. All translations of titles and terms are mine (JR)
2 The chapter is, however, not entitled as an introduction or preface
3 Their contributions are duly noted next to the paper titles.

constraints of the cognitive and phonetic (articulatory and perceptual) levels. These three points form the base of the papers in the volume reviewed here.

Section 1 Theoretical models This part comprises three chapters:

1. Daniel Recasens, "Adaptation of place of articulation in consonant groups of Catalan in light of the DAC Model," (translated by Gabrielle Konopczynski, pp. 19-35). Following the "Degree of Articulation Constraint Model" (DAC) the aim of this paper is to test whether effects of C2 on C1 reflect articulatory planning to prepare the phoneme target, and the effects of C1 on C2 become more constrained by the instantaneous condition and the bio-mechanics of the articulatory organs. On the one hand, these processes may be phonological – and thus categorical; on the other the C1 effects on C2 can be coarticulatory which would yield variation according to flow, speaker or item. The speakers spoke Majorcan, Valencian and Eastern Catalan while wearing an artificial palate and reading aloud sentences which included consonant clusters with C1 /t, n, l/ followed by C2 /s, r, ʃ/ and *vice versa*. The results confirm that effects of C2 on C1 and C1 on C2 reveal different properties: C2 on C1 effects are associated with phenomenes of articulatory planning, producing regressive assimilation, conditioned C2 is sufficiently constrained and C1 sufficiently unconstrained. C1 on C2 effects are associated with inertia phenomena, applied when C1 is constrained and C2 is not constrained, yielding partial assimilation and variable results. The systematic results for C2 effects on C1 vs. the C1 effects on C2 can be attributed to the fact that phonemes at the beginning of a syllable are stronger than phonemes at syllable ends.

2. Christian Abry, Aude Noiray, Marie-Agnès Cathiard and Lucie Menard, "Movement Expansion Model: a universal, individual and developmental anticipation model" (pp. 37-61). This chapter examines the Model of Movement Expansion (MEM) (Abry and Lallouache, 1995), which tests anticipatory lip expansion due to different vowels in the speech sequence. Due to differences found in the literature between English and French speakers' productions, the research questions whether anticipation is universal or language-specific, and whether the involved lip protrusion is a developing process. These questions were studied with 4 Quebec French and 4 American English speakers as subjects, and with seven 3-8 years old French speaking children. Additional questions concerning the children considered their similarity to adults' behavior and whether there is a critical period for acquiring coarticulatory anticipation. The subjects were studied pronouncing (with and without a bite-block) nonsense sentences with sequences of i-C-u to i-CCCCC-u, (C = /s, k, t/). The study analyzed lip movement (protrusion, hold and contraction) synchronized with voice recordings using audio-visual and cinematic recording. The results revealed differences between the French and English speakers. The findings of the developmental study in children showed that anticipatory coarticulation was acquired by age 5, and with some evidence of it already by age 3:6. Thus, the period between 3 to 6 years seemed critical for acquiring anticipation, with children learning the adult model. These findings corroborate that MEM can be used beyond French with English speakers.

3. René Carré, “Coarticulation in speech production: Acoustical aspects,” (pp. 63-74). This article also investigates previous theoretical models, basing the analysis on DRM (Distinctive Regions Model). Using DRM enables the author to distinguish between vocal tract dependent features and speech sequence (coarticulation) features. He defines three phases in this process: where the vocal and consonantal gestures are synchronic and when either of them precedes the other. He also shows that direction and speed of movement in speech (shown by F1 and F2 transitions in diphthongs) and not only the form of the vocal tract are important processes for characterizing the following vowel. This distinction was also tested percpetually and the author suggests that perception should begin with phonological rather than with phonetic (vowel triangle) aspects, and that transition speed be considered an intrinsic vowel feature rather than formants, which he considers to be extrinsic.

Section 2 Observation and Instrumentation Methods This part includes the following four chapters:

4. Véronique Delvaux and Bernard Harmegnies, “Contextual nasality and vowel opening in French: an aerodynamic study” (pp. 77-90). This paper examines the relationship between nasality and (open, closed) vowel types, using aerodynamic calculations of the air flow at the nose and mouth and the total differences between them. All of the French vowels were studied, some following a nasal consonant (nasal context), some preceding it, and some without a nasal context. The subjects were 8 Belgian speakers of French as their mother tongue. The results of this investigation show that the nasalization effect is most important for the high vowels /i, y, u/. The paper is also important because it reports about all of the French vowels in a systematic manner, which enables comparisosn of results with previous studies which have dealt with the subject only partly.

5. Mohamed Yeou, Kiyoshi Honda, Shinji Maeda and Mohamed Embarki, “Laryngeal adjustments during consonant sequence production in Moroccan Arabic” (pp. 91-100) Consonant sequence processes are studied here using the speech of a Moroccan Arabic speaker. The chosen consonants were /x, s, sˤ/ followed by /k/ or /t/, as well as the geminates /ss/ and /sˤsˤ/. All were studied as part of one word and with word juncture intervening in the sequence. A Palatoglottograph (PGG) was used to record the speech utterances, which enabled non-invasive recording of laryngeal and vocal tract motions. The analysis was aided by a program written by Maeda. The paper discusses the results of word limits and speech flow effects. The laryngeal-oral coordination, word limits and speech flow were found to affect the laryngeal-oral coordination. The suggested explanation is that air flow during fricatives and voiceless plosives is mainly controlled by the oral constriction, with weak demands on the glottis. In contrast, the plosives, mainly characterized by aspiration noise, are required to control the vocal cords very quickly after the occlusion release.

6. Paula Marins, Inês Carbone, Augusto Silva and Antonio Teixeira, “Coarticulatory effects on European Portuguese: A first MRI study” (pp. 101-115, translated from English by Philippe Boula de Mareüil). According to the authors,

this is the first study of Portuguese coarticulation using MRI. Therefore, this method which has its advantages and shortcomings (as everything) is described at some length. The study examines the unvoiced and voiced fricative and stop consonants (/f, s, ʃ, v, z, ʒ, p, t, k, b, d, g/) in the VCV environment /a, i, u/. Coarticulation effects are described and presented in several figures, showing the differences between the cases of the vowel environments for the various consonants. The results are in line with previous research (e.g., Farnetani, 1999), where the fricatives are most resistant to coarticulation, while the least resistant are the stops. Here, the only consonants that did not show coarticulation effects and were most resistant to coarticulation were /ʃ, ʒ/.

7. Galina Kedrova, Nikolai Anisimov, Leonid Zaharov and Yuri Pirogov, "Articulatory patterns in anticipatory pauses in Russian: an MRI investigation" (pp. 117-129, translated from English by Christelle Dodane). As in the previous paper, these authors use MRI for studying coarticulation. Here, the six cardinal vowels of Russian are examined as part of a larger project about anticipatory coarticulation features during pauses between and after phonation. A survey of the few studies on this topic in Russian (e.g., Skaluzob, 1979) is followed by description of their own work. The findings support the model that in Russian, anticipatory coarticulation is a systematic process vs. the model that coarticulation is mainly operated by physiological, mechanical and inertial forces. The behavior of anticipatory coarticulation depends very much on the anticipated vowel and is thus partly phonology-dependent. Three categories of vowel weight could be suggested in the vocal system of the subjects due to the functioning of the position of the back of the tongue inside a global vocal tract contour of patterns of anticipatory coarticulation. This motor stereotype could be considered the main factor operating in the subjects' vocal articulatory activity. In addition, three tongue height levels can be considered most important for grouping pre-adjustment articulatory patterns.

Section 3 Acoustic signs This section is comprised of two papers:

8. Cedric Gendrot and Martine Adda-Decker, "Influence of consonantal context and vowel duration on oral vowels centralization in French" (p. 133-142). The paper investigates the theoretical problem of whether formant centralization effects are due to vowel contexts or vowel durations. The authors used automatically acquired data from 25 hours of recorded natural French speech and analyzed it by systematic grouping the data into four consonant contexts (labial, alveolar, palato-velar, and uvular) vs. all the French oral vowels in the context of C1VC2 syllables. The results show that vowel centralization occurs in about 43% of the different contexts, and that the alveolar context is the most centralizing factor (in 55% of the syllables). However, the duration effect on F1/F2 centralization appear stronger than any context.

9. Mohamed Embarki, Christian Guilleminot, Mohamed Yeou and Sallal AlMaqtari, "Coarticulatory effects of pharyngeal consonants in VCV sequences in Modern Arabic and Dialectal Arabic" (p. 143-154). Another aspect of coarticulation relates to phenomena due to pharyngeal/vowel adjacency. This feature is studied in

the linguistic systems of Modern Arabic (MA, the inheritor of Classical Arabic, still related to its phonological system, though acquired at school and not used as a native tongue) and an Arabic dialect (DA). Sixteen native speakers of Arabic[4] pronounced VCV words within a carrier sentence (where C= /t, d, s, ð/ and /$t^ʕ$, $d^ʕ$, $s^ʕ$, $ð^ʕ$/; V= /i, u, a/). F1, F2 values and differences among them were measured for all the utterances, using Praat. The results show coarticulation differences between MA and DA in vowel offset and onset (in the VCV context), and different effects between various vowels and consonants. Coarticulation differences are stronger for the pharyngeal than for the non-pharyngeal consonants. In particular, /i/ resists coarticulation force; but in general, vowel resistance to coarticulation force is weaker in a non-pharyngeal environment than in a pharyngeal one. Coarticulation effects are slightly higher in MA than in DA (when compared to existing literature). These results are in line with the authors' hypotheses.

Section 4 Perception Here we also find two papers:

10. Patrice Speeter Beddor, "Perception of variation due to coarticulation" (pp. 157-176, translated from English by Véronique Delvaux). This paper examines listeners' various perceptual judgments (behavior) in various cases of coarticulation. Lindblom's (1990) theory predicts, for example, that listeners avoid considering coarticulation in order not to hamper spoken utterance understanding. Other literature suggests that listeners accommodate or compensate for coarticulation effects without this making speech identification more difficult for them. Beddor then presents the results of his work on nasal vowels (ṽ, CṽC, NṽN) contexts with English speakers. Nasalized vowel coarticulation compensates at least partly in several environments, but listeners do not systematically hear whether a vowel is nasalized or not in NṽN or CṽC environments. Other tests show that listeners use nasality due to coarticulation to correct perceived vowel height and that coarticulation due to nasality (N) is equal for them to its acoustic effects on vowels (ṽ). Such results have implications in listener-directed speech production (hyper-coarticulation) and coarticulation phonology (coarticulation effects in different languages).

11. Michael Grosvald and David Corina, "Exploring the limits of long-distance coarticulation V-to-V" (pp. 177-186, translated from English by Christelle Dodane). This study explores how far coarticulation effects can reach and be sensed in the speech flow. Since such effects were not found for all studied subjects (reported in the literature) this paper reports a study with 20 English speakers' recorded sentences, where the target vowels preceded the context vowels (at the end of the sentence) by 1, 2 and 3 syllables. While two subjects revealed significant results for all three conditions (one of whom had > 300 ms interval between the first target and the context vowels), the others had only marginally significant results. The authors

[4] The exact origin of the speakers, i.e., the dialect, is not noted.

describe additional perception tests raising more questions about individuals' V-to-V coarticulation perception/production skills.

Section 5 Phonological representation and linguistic constraints This last part has four chapters:

12. Edward Flemming "Coarticulation grammar" (pp. 189-211 translated from English by Christelle Dodane and Mohamed Embarki). Analyzing a few previous theoretical models, this paper claims that coarticulation operates as part of the phonetic/phonological system of a language, and is constrained by the same constraints. Coarticulation processes, compared in four languages (English, French, German and Hindi), serve to justify this approach. Flemming analyzes two cases: tonal coarticulation and coarticulation in coronals-dependent vowel fronting. Flemming's method uses an optimality theory approach with phonetic/phonological features. To calculate coarticulation effects, he uses weights related to the F2 of the vowel (target) and consonant (locus) of the examined movement. Differences between the languages are discussed in terms of difference systems (e.g., existence of /u, y/ in the system or just /u/) and vowel duration (German /u/ is longer than in the other examined languages). Thus, Flemming argues that explaining phonology by phonetics enables him to explain the phonetic patterns themselves. Coarticulation analysis clarifies the nature of the influence of mechanical properties of speech physiology on the linguistic sonorant patterns, which vary among languages.

13. Hyunsoon Kim, "Palatalization in the Korean language as an example of coarticulation: Ciné-MRI, stroboscopic and acoustic data of gradual tongue motions" (pp. 213-226; translated from English by Jalal-Eddin Al-Tamimi). Palatalization (of /t/ > /ts/ etc.) occurs in Korean within word boundaries, as well as when a morpheme is attached to the word. An MRI experiment which follows tongue movements (raising and fronting) during speech and acoustic measurements of the same recorded words (spoken by 10 subjects) shows that palatalization of coronal consonants (/t, t^h, t', ts, ts^h, ts'/ in the environment of /i/ is a phonetic rather than a phonological process, and thus is part of anticipatory coarticulation processes in this language.

14. Chakir Zeroual, Philip Hoole and John Esling, "Articulatory and acoustico-perceptive constraints related to the production of emphatic /k/ in Moroccan Arabic" (pp. 227-240). Emphatization (pharyngalization or velarization) in Arabic has attracted much scholarly attention. The number and quality of emphatic consonants varies in Arabic dialects (and differs from Classical Arabic phonology). This paper is the first study that focuses on the pronunciation of /k/ compared to /q/ in the context of coarticulation.. Its participants speak Moroccan Arabic. The study examines coarticulatory effects that appear (and prevent) the occurrence of an emphatic (e.g., /$t^ɕ$/) in the adjacency of /k/. The authors claim that such a sequence as */$t^ɕ$k/ is prevented by the different tongue directions and movements involved in

each consonant. The investigation was conducted by naso-endoscopic observation[5] of the back of the vocal tract and analysis of the acoustic values of F1, F2 in initial and median /a_2/, /i_2/ sequences of /-$i_1t^{\varsigma}i_2$/ and /-$a_1t^{\varsigma}a_2$-/ in several word and non-word series. The recorded formant values and observations showed that, as initially assumed to be an articulatory constraint, tongue back was fronted and higher in /k/ than in /t^{ς}/. The fact that emphatic /k/ would approach /q/ is prevented, probably due to acoustic-perceptive constraints. The authors therefore suggest conducting perceptual studies of the relative weight of these two constraints.

15. Stephanie Lain, "Articulation force and [voicing] in VCV coarticulation: data of two dialects of Venezuelan Spanish" (pp. 241-258, translated from English by Christelle Dodane). After presenting theories of the structure and changes involved in voicing in several languages, and describing this subject in Venezuelan Spanish, the goal is presented: the study of voicing in a low land dialect and a mountainous area dialect (Margarita and Mérida) in this country. The paper examines recordings of 10 speakers (5x2) uttering the Spanish voiced and unvoiced consonants and vowels in tri-syllabic $C_1V_1rV_1C_1V_1$ sequences (e.g., /doróðo/), stressing the penultimate syllable. Formants and duration values are measured and differences between the two chosen dialects are compared. Inter-dialect and inter-speaker differences are found significant for only some of the measures, and these findings are to be further considered for number of speakers, different sexes, and dialects. Approximants instead of stops also occurred. The author plans to study all these aspects perceptually.

In sum, this is an important up-to-date collection on coarticulation for French-reading phoneticians and linguists (in spite of minor typo and translation errors). The book presents many theoretical and experimental aspects of coarticulation in vowels, consonants, and vowels + consonants, and shows the relevance of coarticulation to other phonetic areas. The various aspects are investigated using sophisticated technological tools (e.g., for MRI, articulography and naso-endoscopy), combined with computerized acoustic analysis methods (e.g., Praat or WaveSurfer) and statistical calculations (using, Praat or SPSS). This wealth of analysis techniques reveals interesting features of coarticulation in different languages and raises additional questions for future research.

Reference

Abry, C. and T. Lallouache 1995. Le MEM: Un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français. *Bulletin de la Communication Parlée* 3: 85-99.

Bladon, R.A.W. and Al-Bamerni, A. 1976. Coarticualtion resistance in English /l/. *Journal of Phonetics* 4: 137-150

Farnetani, E. 1999. Coarticulation and connected speech processes. In: Hardcastle, W. J. and Laver, J. (eds.): *Handbook of Phonetic Sciences.* Oxford: Blackwell, 371-404.

[5] Zeroual et al. thank and acknowledge Dr. Lise Crevier-Buchman's work in this part.

Hammarberg. R. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4: 353-363.
Hardcastle, W. J. and Hewlet, N. (eds.) 1999. *Coarticulation: Theory, Data and Techniques.* Cambridge UK: Cambridge University Press.
Kuehnert, B. and Nolan, F. 1999. *The origin of coarticulation.* In: Hardcastle, W. J. and Hewlet, N. (eds.) 1999. *Coarticulation. Theory, Data and Techniques.* Cambridge UK: Cambridge University Press, 7-30.
Lindblom, B. 1990. Explaining phonetic variation: a sketch of the HandH theory. In: Hardcastle, W.J. and Marchal, A. (eds.): *Speech Production and Speech Modeling.* Dordrecht: Kluwer Academic, 403-439.
Perkel, J.S. and Matthies, L.M. 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within and cross-subject variability. *Journal of the Acoustical Society of America* 91(5): 2911-2925.
Skaluzob, L. 1979. *Dinamika zvukoobrazovanija (po dannym kinorentgenografi-rovanija*. Kiev: Naukova Dumka.

B. Elan Dresher (2009):
The Contrastive Hierarchy in Phonology

Cambridge University Press, Cambridge, 302 pp.
Hardback and paperback, ISBN 9780521889735. Hardback: 68£

Reviewed by: **Noam Faust**
The Hebrew University, Jerusalem, Israel
e-mail: faustista@yahoo.com

Introduction: B. Elan Dresher's book, *The Contrastive Hierarchy in Phonology*, achieves two goals. First, as its title suggests, it argues that contrastive phonological features are best arrived at through a hierarchical conception of their arrangement. The second goal, which is achieved in the process of arguing for the author's view, is the thorough presentation of the evolution of the idea of contrast in phonological theory. Although the explicit goal is the first one, the second occupies at least half of the conceptual and actual content of the book. It seems, moreover, that the strength of this book lies much more in the critical scrutiny of past studies than in arguing for its own thesis. Be that as it may, the output is undoubtedly not only the first of its kind, but also the fruit of considerable scholarly work. It is therefore a major reference for anyone interested in formalizing contrast in phonology and in the question of how linguistic knowledge is structured in that respect. In what follows, I review the book chapter by chapter and conclude with my own general remarks.

Contents by chapter: The first chapter (pp. 1-10) introduces the basic theoretical problem of the book: linguistic systems with identical (or very similar) speech sound arrays behave differently with respect to these sounds. This differentiating behavior cannot be attributed to phonetics, because the sounds are phonetically identical. The structuralist analysis of this fact was to organize the sounds in question differently in the conceptual space of the two systems. Ever since these first studies, contrast was a key notion in this respect (although not explicitly): the sounds were presented as

contrasting with each other or not. The different organizations accounted for the differences in sound patterns.

The second chapter (pp. 11-36) is maybe the most significant in this book in terms of theoretical import. It discusses the formalization of contrast: how does one determine what aspects of sound - what features and feature values - are contrastive in a given language? One way of doing this is what the author calls the Pairwise Algorithm: those features are deemed contrastive that serve to distinguish between *pairs* of phonemes. Another way of arriving at a contrastive set is through Feature Ordering. In this strategy, some distinctions are more primary than others, so that first, all of the sounds are distinguished with respect to one feature, and only then distinctions are made with respect to features that are "lower" in the ordering. This is the Contrastive Hierarchy in the title of this book. Crucially, the two approaches do not yield the same results, so empirical and conceptual adequacy can be examined. Moreover, Feature Ordering is more malleable: one may say that different languages arrange their features in different orders, thereby beginning to solve the motivating problem for the entire book (i.e. similar arrays, different behavior).

The rest of the chapter is devoted to showing that on the conceptual level, the Pairwise Algorithm is considerably inferior to Feature Ordering. Indeed, the author further shows that the Pairwise Algorithm must presuppose an ordering of features in order to be applied.

Chapter 3 (pp. 37-75) continues to illustrate the advantages of Feature Hierarchies, although from a more empirical (yet still very idealized) point of view. It surveys the work of the first structuralists and shows that contrast was a basic notion in their analyses of sound systems. However, interestingly, the approach to contrast was never made explicit by the proponents themselves, so that further interpretation is needed in order to unveil the method adopted in each study. Furthermore, it is shown that in these early and fundamental studies, empirical considerations (as opposed to conceptual ones) favored Feature Ordering over the Pairwise Algorithm. The most important question that is raised in the process of this critical survey is how to arrive at contrastive specifications. The answer, on which the author cites D.C. Hall (2007), is termed The Contrastivist Hypothesis: "The phonological component of a language L operates only on those features which are necessary to distinguish the phonemes of L from one another." For instance, if a language exhibits voicing assimilation, then voicing must be contrastive in this language.

The fourth chapter (pp. 76-103) continues in the same spirit. It surveys works from the 50's and 60's and the influential work by Jakobson and Halle. This is maybe the most historically-oriented chapter in the book. Its main aim is to show how Feature Ordering first continued to be an implicit practice, but then was eventually dissociated from the Contrastivist Hypothesis and largely abandoned and replaced by minimality and economy considerations. This chapter also introduces the problem that will be referred to in the rest of the book, namely regressive

assimilation in Russian, and the theoretical discussion around it. The chapter concludes with an interesting summary of Stanley's (1967) arguments against the analysis in Halle (1959), around the use of zeros in phonological representation.

Chapter 5 (pp. 103-137) treats the advent of early Generative Phonology and the fate of the Contrastivist Hypothesis in it. This hypothesis, the author shows, was rejected in the *Sound Patterns of English* and subsequent generative work as a part of the general rejection of earlier structuralist practice. If so, how did this new current of phonological thought derive feature specifications? Three alternatives are discussed in the chapter: Markedness theory, Underspecification, and "theories of feature organization". The common denominator of these alternatives, which is not shared by the author's proposed hierarchy, is the attempt to arrive at universal statements about feature organization. Other aspects of these theories are shown to be actually shared by the Contrastive Hierarchy.

Chapter 6 (pp. 138-161) looks at contrast in Optimality Theory (OT) and how contrast has been formalized within it. After surveying past analyses, the author provides his own version of how this should be done, stating that the Feature Hierarchy is not incompatible with OT. See Scheer (2010) for criticism on this analysis.

Chapter 7 (pp. 162-210) is the most empirically-oriented chapter in the book. Readers who are interested more in the application of the Contrastive Hierarchy should start by reading this chapter (although it does often refer to previous chapters). In this chapter, it is shown that the hierarchical organization of features can account for differences between related languages (see especially the discussion of metaphony in Romance dialects), or between different stages of one language (vowel harmony in Classical and Modern Manchu). The chapter also includes a relatively in-depth account of loanword adaptation in terms of hierarchy: the analysis accounts for the different adaptation strategies in Hawaiian and Maori, which have very similar phoneme inventories. Another interesting discussion concerns the implementation of the Contrastive Hierarchy approach to acquisition. Finally, there is a short discussion of how counter-examples may be treated, citing again D.C. Hall's (2007) proposal of "prophylactic features", i.e. features that are present in the representation but for some reason ignored by the phonology. This disturbing notion is however said to represent "a minimal retreat from the Contrastivist Hypothesis" (p. 209). For further discussion of the topic of this chapter see my remarks below, and the more elaborate points in Scheer (2010).

The eighth chapter of the book (pp. 211-249) is a comparison to other modern approaches to contrast. The most elaborate discussion in the chapter concerns East Slavic post-velar fronting from both a diachronic and synchronic perspective, contrasting the author's approach with Padgett's Dispersion Theory (Padgett, 2003, 2004). Other comparisons are with Structures Specification theory (Frisch et al. 2004) and Visibility theory (Calabrese, 2005). When the author does not point at wrong predictions made by the competing theories, he claims that the same correct

generalizations can be arrived at using the Contrastive Hierarchy in a more intuitive, elegant manner.

Chapter 9 (p. 250) is a short conclusion.

Reviewer's remarks:

I think this book is a worthy scholarly effort and a welcome contribution to the study of the topic. I do however have several reservations, of which I will now mention three.

First, a practical point: I found the book relatively hard to read, maybe because it is at once rich in data sources and abstract concepts, but poor in thorough examinations of the specific phenomena, which could have further explained the theoretical tools. Many of the relevant cases are presented with two or three examples, a "wrong analysis" and a hierarchy that fits the phenomena, and the argumentation moves on to the next set. The amount of data and diverse phenomena to take into consideration could have been made an easier task to tackle by more elaborate presentations of each of the phenomena.

The second point is related to the first, although it is more theoretical. The correctness of an analysis, it seems to me, does not rely solely on its success in covering (or "accounting for") the facts that it set out to cover. The analysis has to be shown to make correct predictions elsewhere in the system. There are very few cases - if any - in this book where this kind of reasoning is explored. Instead, the correctness of the Contrastive Hierarchy is deduced from the failings of other theories in covering what the Hierarchy can cover, rather than from the confirmation - or admitted lack thereof - of the Hierarchy's predictions. [6]

The third point is also methodological. Theories have to be constrained. If I am not mistaken, the Contrastive Hierarchy predicts that all possible feature orderings are possible. Does one find all possible hierarchies in the world's languages? I guess that the answer is "no"; but if so, why? How can the Contrastive Hierarchy be constrained so that it doesn't predict that? An explicit discussion of this topic (and similar ones in the same spirit) is certainly missing from the book.

Despite these reservations, the fact remains that this book is certainly an important book for anyone interested in contrast in phonological thought. It is important for its critical and helpfully interpretive presentation of this aspect in past studies. It is also important because it brings forth an implicit practice and formalizes it, so that it is clear that the Contrastive Hierarchy is at least a competitor in the analysis of contrast in phonological systems. In this respect, the book is also important in that the richness of the phenomena to which the Hierarchy is applied exemplifies several of its strong points. Finally, it is an important book because it spells out assumptions about contrast that have previously been only implicit, such as the above mentioned Contrastivist Hypothesis.

[6] To see how such a discussion would look, see Scheer's (2010) review of the book.

References
Calabrese, A. 2005. *Markedness and economy in a derivational model of phonology*. Berlin: Mouton de Gruyter.
Frisch, S. A,, Pierrehumbert, J. B., and Broe, M. B. 2004. Similarity avoidance and the OCP, *NLLT* 22: 179-228.
Hall, D. C. 2007. *The role and representation of contrast in phonological theory.* Ph.D. dissertation, Department of Linguistics, University of Toronto
Halle, M. 1959. *The sound pattern of Russian: a linguistic and acoustical investigation.* The Hague: Mouton. Second printing, 1971
Padgett, J. 2003. Contrast and post-velar fronting in Russian. *NLLT* 21.1: 39-87.
Padgett, J. 2004. *Russian vowel reduction and Dispersion Theory. Phonological Studies* 7, Kaiakusha, Tokyo, pp. 81-96.
Scheer, T. 2010. How to marry (structuralist) contrast and (generative) processing (review of Dresher 2009, The Contrastive Hierarchy in Phonology). *Lingua* 120: 2522-2534.
Stanley, R. 1967. Redundancy rules in phonology. *Language* 43: 393–436.

Dagmar Barth-Weingarten, Elisabeth Reber and Margret Selting (eds.) (2010):
Prosody in Interaction

(Studies in discourse and Grammar Series): John Benjamins Publishing Company, Amsterdam/Philadelphia, XXI+406 pp.,
Hardback, ISBN**:** 978-90-272 2633 4, Price: Euro 99.0 / US$ 149.-

Reviewed by: **Judith Rosenhouse**
SWANTECH Ltd. Haifa, 32684 Israel
e-mail: swantech@013.net

This volume is based on the conference "Prosody and interaction" which was held at the University of Potsdam, Germany, in September 2008. It is also dedicated to Elizabeth Couper-Kuhlen, on the occasion of her 65th birthday and her outstanding academic achievements.

The volume has four parts. The first part is an Introduction with two chapters: Margret Selting's "Prosody in interaction: State of the art" (3-40) and Arnulf Depperman's "Future prospects of research on prosody: The need for publicly available corpora: Comments on Margret Selting "Prosody in interaction: State of the art"" (41-47). This pattern – a chapter followed by a commenting paper – continues throughout, although not with all the chapters.

Part I is entitled "Prosody and other levels of linguistic organization in interaction." This part contains seven chapters:

"The phonetic constitution of a turn-holding practice: Rush-throughs in English talk-in-interaction" by Gareth Walker (51-72); then Susanne Günthner's comments on Walker's paper entitled "Rush-through as social actions: comments on Gareth Walker "The phonetic constitution of a turn-holding practice: Rush-throughs in English talk-in-interaction" (73-79).

Next comes “Prosodic constructions in making complaints” by Richard Ogden (81-103) which is followed by Auli Hakulinen’s “The relevance of context to the performing of a complaint: comments on Richard Ogden “Prosodic constructions in making complaints”” (105-108).

The next two papers are by Geoffrey Raymond: “Prosodic variation in responses: The case of type-conforming responses to yes/no interrogatives” (109-129); and John Local, Peter Auer and Paul Drew “Retrieving, redoing and resuscitating turns in conversation” (131-159). The last paper in this section is by Harrie Mazeland and Leendert Plug “Doing confirmation with ja/nee hoor: Sequential and prosodic characteristics of a Dutch discourse particle” (161-188). These three papers are not accompanied by comments.

“Part II. Prosodic units as a structuring device in interactions” contains six chapters:
”Prosodic constructions in making complaints: A participant’s category?” by Beatrice Szczepek Reed (191-212) is followed by Jan Anward’s “Making units: Comments on Beatrice Szczepek Reed Prosodic constructions in making complaints: A participant’s category?” (213-216).

The next paper is by Friederike Kern “Speaking dramatically: The prosody of live radio commentary of football matches” (217-237) and the comments on this chapter by Johannes Wagner, entitled “Commentating fictive and real sports: comments on Friederike Kern “Speaking dramatically: The prosody of live radio commentary of football matches” (239-241).

This part ends with Bill Wells’ “Tonal repetition and tonal contrast in English care-child interaction” (243-262) and the comments on this chapter by Traci Walker “Repetition and contrast across action sequences: comments on Bill Wells “Tonal repetition and tonal contrast in English care-child interaction” (263-266).

“Part III Prosody and other semiotic resources in interaction” has eight chapters: Elisabeth Gülich and Katrin Lindemann write about “Communicating emotion in doctor-patient interaction: a multidimensional single-case analysis” (269-294). This is followed by Elisabeth Reber’s paper “Double function of prosody: processes of meaning-making in narrative reconstructions of epileptic seizures: Comments on Elisabeth Gülich and Katrin Lindemann “Communicating emotion in doctor-patient interaction: A multidimensional single-case analysis” (295-301).

Hiroko Tanaka’s paper “Multidmodal expressivity of the Japanese response particle *Huun*: displaying involvement without topical engagement” (303-332) is next and it is followed by Dagmar Barth-Weingarten’s “Response tokens – A multimodal approach: Comments on Hiroko Tanaka’s “Multidmodal expressivity of the Japanese response particle *Huun*: displaying involvement without topical engagement” (333-338).

Cecilia E. Ford and Barbara A. Fox “Multiple practices for constructing laughables” (339-368) and the comments paper by Karin Birkner “Multimodal laughing: Comments on Cecilia E. Ford and Barbara A. Fox “Multiple Practices for constructing laughables” (369-372).

The last pair of papers is by Charles Goodwin "Constructing meaning through prosody in aphasia" (373-394) and Helga Kotthoff's "Further perspectives on cooperative semiosis: Comments on Charles Goodwin "Constructing meaning through prosody in aphasia" (395-399).

The Series Editor, Sandra A. Thompson, mentions in her Foreword, that this volume shows the high set of standards which Elizabeth Couper-Kuhlen in fact set forth in her book *Prosody in Conversation* (1996). The volume also reflects the great progress made in this subfield since then. The above list of papers demonstrates the variety of topics discussed in this volume, which combine linguistic and conversational analysis aspects while focusing on the role of prosody in social interaction. The editors of this book, too, point at this direction (in the Preface) and stress that this field combines established linguistic studies of prosody in action. It likewise reveals the expanded scope of this field of research by studies of the situation in new languages and language elements. In addition, the book offers readers the possibility to see and/or hear online numerous examples of the discussed topics through video clips (.mov) and audio files (.wav) at http://dx.doi.org/10.1075/sidag.23.media. Those examples are marked at appropriate points in the papers. All of the chapters include examples of relevant speech utterances or conversation parts, using various systems of transcription/analysis. Sometimes, intonation traces are added to the examples. Several chapters include speech spectrograms and prosody (pitch) traces using several analysis systems (the papers by Walker, Ogden, Local et al., Mazeland and Plug, B. Szczepek Reed, F. Kern, H. Tanaka, Ford and Fox), photographed pictures (Tanaka, Ford and Fox) and drawn (illustrated) figures. These devices indeed reflect the technical progress of this field, which helps analyze and understand the issues involved in this area. Each paper is also accompanied by a comprehensive bibliographic list, which is always an important contribution for the readers.

Though this subfield of research is relatively young, this volume brings "the state of the art" home to readers with cases analyzing various speech-turn structures (Local et al.), doctor-patient interaction (Gülich and Lindemann), aphasic communication, (Goodwin), and laughter (Ford and Fox), to mention only a few issues, which particularly attracted this writer's attention. Selting's introductory chapter is a thorough study of the field organized by six questions. These questions sum up prosody in general, the importance of studying it, its importance for phonology/phonetic studies, researchers of prosody in interaction, as well as the current and future research questions and the challenges involved in this study. Her main conclusion of this survey is that "[W]hile we can always look at the details of prosody and other forms of utterances as an intermediate step, we always in the end need to come back to the systematic analysis of actions and sequences in interactions and integrate our more form-related analyses into this" (p. 30). In sum, this is a very interesting book on the application of prosody in live communication and is a good read for both students and researchers of the field.

Sharynne McLeod and Brian A. Goldstein (eds.) (2012):
Multilingual Aspects of Speech Sound Disorders in Children
(Communication Disorders across Languages Series)
Multilingual Publishing, Bristol, Buffalo, Toronto, XXIX + 289 pp.,
Paperback, ISBN: 978-1847 695 123, Price: £ 29.95

Reviewed by: **Judith Rosenhouse**
SWANTECH Ltd. Haifa, 32684 Israel
e-mail: swantech@013.net

This book is the first collection of papers dedicated to multilingual children with speech disorders. This collection reflects the collaboration of 44 authors from 16 different countries reporting about multilingual children's language acquisition problems in 116 languages and dialects from all the continents. We will not list here all the languages and dialects that are discussed in the book, but let us note that they include Icelandic, Cuban Spanish, and Australian indigenous languages, which are rarely studied, as well as bimodal[7] problems of hearing impaired children.

These reports are gathered in 30 papers which are organized in three parts which are: 1. Foundations; 2. Multilingual speech acquisition; and 3. Speech-language pathology practice. Some of the papers describe research and analyze linguistic and socio-linguistic issues involved in the different language environments of the multilingual children, while others focus on the speech and language specialists' work in the field. These latter papers are usually much shorter than the former group. In this manner the book contains both theoretical and pragmatic material.

The term "multilingualism" is used in this book very broadly, and thus children in this book may be weak or strong bilingual/multilingual speakers.[8] In addition, though many languages are discussed, the papers describe bilingual children, and not tri-lingual or more, which would reflect "real" multilingualism. The book focuses on speech sounds, which are basic, and relatively early, building blocks of any language. The pictures presented for the different languages differ since languages and cultures have different speech sound systems and social habits of speaking (e.g. in some cultures, children speak only when addressed by adults, and even then concisely).

In addition, the thirty chapters of the book discuss not only different languages and dialects and their interactions but also many relevant linguistic topics for research in this field. Ingram devotes most of his chapter to analyzing the issue of

[7] K. Crow, the author of that paper prefer this term to "bilingual." Due to sign language structure.

[8] This term is in line with authors such as Valdés and Figueroa (1994) who consider bilingualism a relative condition (see Preface, p. XXVII)

speech and acquisition errors in different languages, viewing them rather as phonological rather than as phonetic-articulatory errors, i.e., not only articulation-complexity-dependent but also phonetic frequency and functional load dependent. His main point is that it is definitely important for phoneticians and speech clinicians/therapists to learn about language acquisition features of more than the single language they usually work with. The other papers examine sociolinguistic and cultural aspects, transcription problems and methods, assessment methods of children's articulation and tools for this task, acquisition stages from babbling to child and adult, speech analysis, acoustic features, prosody, perception, intervention methods, and related literacy effects. Thus, the full list of papers in this book is as follows:

"Part 1 Foundations" includes the following papers:

1. David Ingram: "Prologue: Cross-linguistic and multilingual aspects of speech sound disorders in children" (3-12).

2. Madalena Cruz-Ferreira: "Sociolinguistic and cultural considerations when working with multilingual children" (13-23)

3. Carol Stow, Sean Pert and Ghada Khattab: "Translation to Practice: Sociolinguistic and cultural considerations when working with the Pakistani heritage community in England, UK (24-27)

4. Cori J. Williams: "Translation to Practice: Sociolinguistic and cultural considerations when working with indigenous children in Australia" (28-31)

5. Martin J. Ball: "Vowels and consonants of the world's languages" (32-47)

6. Sue Peppé: "Prosody in the world's languages" (42-52)

7. Sue Peppé, Marine Coene, Isabelle Hesling, Pastora Martinez-Castilla and Inger Moen: "Translation to practice: Prosody in five European languages" (53-36)

8. Susan Rvachew, Karen Mattock, Meghan Clayards, Pi-Yu Chiang and Francoise Brosseau-Lappe "Perceptual considerations in multilingual adult and child speech acquisition" (57-67)

Part 2 includes the following four papers:

9. Barbara L. Davis and Sophie Kern: "A complexity theory account of Canonical babbling in young children" (71-83)

10. Brian A. Goldstein and Sharynne McLeod: Typical and atypical multilingual speech acquisition (84-100)

11. Karla N. Washington: "Translation to practice: Typical bidialectal speech acquisition in Jamaica" (101-105).

12. Thóra Másdóttir: "Translation to practice: Typical and atypical multilingual speech acquisition in Iceland" (106-110)

Part 3 contains the rest of the papers:

13. Sharynne McLeod: "Multilingual speech assessment" (113-143)

14. Sharynne McLeod: "Translation to practice: Creating sampling tools to assess multilingual children's speech" (144-153)

15. Seyhun Topbaş: "Translation to practice: Assessment of the speech of multilingual children in Turkey" (154-160)

16. Raúl F. Prezas and Raúl Rojas: “Translation to practice: Assessment of the speech of Spanish-English bilingual children in the USA” (161-164)

17. Carol Kit Sum To and Pamela Sau Ping Cheung: “Translation to practice: Assessment of children’s speech sound production in Hong Kong” (165-169)

18. Jan Edwards and Joseph P. Stemberger: “Transcription of the speech of multilingual children with speech sound disorders” (170-181)

19. B. may Bernhardt and Joseph P. Stemberger: “Translation to practice: Transcription of the speech of multilingual children (1820-190)

20. Kathryn Crowe: “Translation to practice: Transcription of the speech and sign of bimodal children with hearing loss” (191-195)

21. Shelley E. Scarpino and Brian A. Goldstein: “Analysis of the speech of multilingual children with speech sound disorders” (196-106)

22. Minjung Kim and Carol Stoel-Gammon: “Translation to practice: Acoustic analysis of the speech of multilingual children in Korea” (207-210)

23. Helen Grech: “Translation to practice: Phonological analysis of the speech of multilingual children in Malta” (211-213)

24. Christina Gildersleeve-Neumann and Brian A. Goldstein: “Intervention for multilingual children with speech sound disorders” (214-227)

25. Annette V. Fox-Boyer: “Translation to practice: Intervention for multilingual children with speech sound disorders in Germany” (228-232)

26. Avivit Ben David: “Translation to practice: Intervention for multilingual Hebrew-speaking children with speech sound disorders in Israel” (233-237)

27. Isabelle Simard: “Translation to practice: Intervention for multilingual children with speech sound disorders in Montréal, Québec, Canada” (238-243)

28. Yvette Hus: “Literacy and metalingusitic considerations of multilingual children with speech sound disorders” (244-256)

29. Ruth Huntley Bahr and Felix Matias”: “Translation to practice: Metalinguistic considerations for Cuban Spanish English bilingual children” (257-262)

30. Brian A. Goldstein and Sharynne McLeod: Multilingual children with speech sound disorders: An Epilogue (263-266)

Being a book about speech and its correct form in contrast with impaired forms, three appendices providing IPA transcription tables (A. normal articulation, revised for 2005, B. symbols for disordered speech, and C. voice quality symbols) appear at the end. These tables are necessary, because the examples in every chapter use transcription. Following the nature of this book, the authors’ names in the list of contributors are written also in IPA transcription, next to their formal spelling.

The numerous chapters of the book present much more information than expected about the linguistic and practical problems in treating multilingual children with speech sound disorders and the possible outcomes of various linguistic interactions. Yet this is the tip of the iceberg, as the authors write, because many questions remain un-answered and unresolved, and lack of knowledge or missing research tools, assessment and treatment are noted in almost every chapter. Thus, the principal readers of this book would be practitioners in speech-language pathology,

therapist, logopedists, etc. But this volume is not less important for students and researchers interested in linguistics, phonetics and/or phonology who will find here a host of interesting facts and topics for further study.

References

Valdés, G. and Figueroa, R.A. 1994. *Bilingualism and Testing: A special Case of Bias*, Norwood, NJ: Ablex.

Walker, Rachel (2011): Vowel Patterns in Language

(Series: Cambridge Studies in Linguistics No. 130)
Cambridge University Press, Cambridge viii + 366 pp.
Hardback: ISBN 9780521513975. Price: £65. Also available as an eBook)

Reviewed by: **Evan-Gary Cohen**
Department of Linguistics, Tel-Aviv University, Israel,
e-mail: evan@post.tau.ac.il

The book investigates the restrictions on vowel patterns in languages, particularly the relationship between various vowels and the positions in which they occur. Chapter 1 (pp. 1-11: "Introduction") presents the two central themes of the book, the effects of word position (positional prominence) on vowel patterning and the relationship between perception and production and vowel pattering restrictions. The book provides a formal model within which the restrictions on which patterns can and cannot occur can be explained.

The prominence-based approach in the book hypothesizes that vowel patterns are largely perceptually driven, although articulatory effects (the enhancement of stressed syllables articulatorily) interact with these perceptual cues. Processing considerations for perception and production also play an important role.

The formal Optimality Theoretical model (Prince and Smolensky, 2004) adopted to deal with the prominence-based systems uses prominence-based licensing constraints. Central to the model is the claim that features are licensed provided that some member of the chain to which the feature belongs is affiliated with a given licensing position. Three types of licensing are discussed: Identity licensing (chapter 6), where licensing for a feature in a non-prominent position is achieved by a duplicated feature in a prominent position; Indirect licensing (chapter 5), where a feature has associations with a prominent position and a non-prominent position; Direct licensing (chapter 7), where a feature is contained wholly within a prominent position.

Chapters 2-4 comprehensively present the fundamentals for the formal analyses provided in 5-8, followed by conclusions in 9.

Chapter 2 (pp. 12-35: "Preliminaries: functional grounding") presents an insightful investigation of the functional grounding for asymmetries in positional prominence and vowel markedness. The roles of perception, production and processing in the formulation of constraints are at the centre of this chapter. Several positions are under focus. Various phenomena related to positional prominence (e.g. resisting lenition or deletion, serving as triggers of vowel harmony, contrast preservation) are discussed. The privileged role of stressed syllables, the special status of initial syllables, evidence of the prominence of final syllables, and stem/root prioritization are all discussed in detail. The roles of specific vocalic features (e.g. Advanced Tongue Root, height) are also dealt with. Predictions for prominence-based licensing phenomena are delineated.

Chapter 3 (pp. 36-63: "Generalized licensing") deals with the formal aspects of the concept of licensing, and is particularly technical in nature. The chapter discusses licensing with respect to prominence, providing a formal model for its expression. An additional topic touched on here is the relationship between morphemes and licensing. For example, in Jaqaru (Aymara; Peru), complete harmony of all features in a preceding stressed syllable is only triggered by certain suffixes (e.g. the triggering suffix /-ni/ 'possessive' vs. the non-triggering suffix /-ni/ 'translocative: 'possessive' - /tʃim**a**-ni/→[tʃim**í**ni] 'with belly' vs. 'translocative' - /mant**a**-ni/→[mant**á**ni] 'to enter towards the speaker').

Chapter 4 (pp. 64-88: "Typological predictions") thoroughly examines typological predictions that were made by prominence-based licensing constraints in conjunction with a set of other constraints that are relevant to a typology that includes licensing-driven assimilation. Starting with disyllables, and then trisyllables and finally long distance effects, 'factorial typology' (i.e., all possible rankings of a given constraint set) is investigated via OTSoft, Version 2.1 (Hayes et al., 2003). The various licensing configurations and their respective constraint interactions are emphasized.

Chapter 5 (pp. 89-144: "Indirect licensing") is the first of three chapters (5, 6, 7) focusing on the description and analysis of vowel patterns involving prominence-based licensing. Each of these three chapters introduces a core constraint ranking structure for patterns under focus. Indirect licensing serves to reduce perceptual difficulty by causing a vowel quality to be produced both in a prominent position and an adjacent non-prominent position (or positions). For example, in Buchan Scots (Germanic; Scotland), high unstressed vowels undergo height assimilation to preceding stressed non-high vowels (e.g., the suffix /-i/: /hér-**i**/→[h**ére**] 'hairy'). The licensing positions discussed in this chapter are the stressed syllable, the initial syllable, and root/stem syllables, which is what one would expect based on phonetic and psycholinguistic strength.

Chapter 6 (pp. 145-192: "Identity licensing") focuses on identity licensing. In such configurations, a vowel quality in a non-prominent position is licensed by a duplicate of the feature in a prominent position. Vowels can also interact at a distance, across transparent material. In Eastern Meadow Mari (Uralic; Russia), for

instance, vowels in word-final syllables assimilate in backness to the vowel in the initial syllable, even if there are intervening vowels, frequently [ə] (e.g. *ém-d**œ*** 'your (pl) medicine' vs. *kutkó-t**a*** 'your (pl) ant'). It is predicted that systems displaying identity licensing will also have forms with indirect and direct licensing configurations.

Chapter 7 (pp. 193-237: "Direct licensing") discusses strict direct licensing. In such cases, perceptual difficulty is minimized by realizing restricted elements only in prominent licensing positions. For example, in Belarusian (East Slavic; Belarus), five vowels contrast in stressed syllables ([i e a o u]), while only three contrast in unstressed syllables ([i a u]), with the mid-vowels /e/ and /o/ lowering to [a]).

Chapter 8 (pp. 238-296: "Maximal licensing") turns to maximal licensing patterns, which are the patterns beyond those driven by prominence-based licensing. Maximal licensing harmony can be triggered by a vowel in a weak position and/or by a vowel which displays some weak property or a combination of properties. First, two patterns are examined where the trigger resides in a strong position, which is the locus of contrast for a particular weak property. The chapter presents a comprehensive case study for the Servigliano dialect (Romance; Italy), which includes two maximal licensing patterns with triggers that are weak by virtue of their properties and/or their position.

In all, Servigliano displays four distinct vowel patterns, each of which shows some sensitivity to relative weakness and/or positional prominence. As such, it provides an excellent testing ground for constraints that drive prominence-based licensing and maximal licensing.

Chapter 9 (pp. 297-313: "Conclusion and final issues") assesses results of this work and highlights some topics that merit attention in future research, such as a similar analysis of consonant systems.

This book presents a well-written, thorough analysis of the role of prominence in vowel patterns, providing further much needed insights into the groundedness of the prominence-based constraint systems within an Optimality Theoretical framework. While much of the book deals with a formal phonological analysis of the various phenomena, phonologists of all levels would find this work extremely useful. The extensive analyses, alongside an abundance of data and in-depth descriptions, allow all linguists researching phonology-phonetic interfaces, as well as linguistic typologists, to benefit from this book.

References

Hayes, B., B. Tesar and Zuraw, K. 2003. *OTSoft 2.1, Software Package*. www.linguistics.ucla.edu/people/hayes/otsoft/.

Prince, A. and Smolensky, P. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

MEETINGS, CONFERENCES AND WORKSHOPS

2013

16–18 January 2013
Variation and Language Processing 2 (VALP2)
Christchurch (New Zealand)
http://www.nzilbb.canterbury.ac.nz/VALP.shtml
ucvalp@gmail.com

16–18 January 2013
The CUNY Conference On The Feature In Phonology And Phonetics
New York (USA)
http://www.cunyphonologyforum.net/featconf.php
feature@cunyphonologyforum.net

16–19 January 2013
10th Old World Conference in Phonology (OCP10)
Istanbul (Turkey)
http://www.ocp10.boun.edu.tr/
ocp10@boun.edu.tr

22–25 January 2013
La Percepción Unimodal y Multimodal del Habla
Madrid (Spain)
http://www.sel.edu.es/?q=node/153

25–27 January 2013
International Conference on Phonetics and Phonology 2013 (ICPP2013)
Tokyo, Japan
http://www.ninjal.ac.jp/phonology/InternationalConference/icpp_2013/home/
phonology@ninjal.ac.jp

28 February - March 1 2013
CROSSLING symposium: Language contacts at the crossroads of disciplines
Joensuu (Finland)
https://wiki.uef.fi/display/CROSSLING/CROSSLING+Symposium+2013
crossling@uef.fi

12–15 March 2013
Prosody and Information Status in Typological Perspective. Workshop at the 35th Annual Meeting of the German Society ofLinguistics (Deutsche Gesellschaft für Sprachwissenschaft, DGfS)
Potsdam (Germany)
stefan.baumannuni-koeln.de or frank.kuegleruni-potsdam.de

18–22 March 2013
Speech in Action
Copenhagen (Denmark)
https://sf.cbs.dk/cphspeech2013
sjusk2013@cbs.dk or exapp@hum.ku.dk

17–18 April 2013
Workshop on Sound Change Actuation

Chicago (USA)
http://lucian.uchicago.edu/blogs/phonlab/sound-change-actuation/

29–30April 2013

International Conference on Language Learning (ICLL 2013)
Johannesburg (South Africa)
https://www.waset.org/conferences/2013/johannesburg/icll/index.php
https://www.waset.org/international.php

4 May 2013

4th Theoretical Phonology Conference (TPC 4)
Taipei, (Taiwan)
http://phonology.nccu.edu.tw/tpc4/

8–10 May 2013

3rd International Conference on English Pronunciation: Issues and Practices (EPIP3)
Murcia (Spain)
https://sites.google.com/site/epip32013/
epip3contact@gmail.com

17–19 May 2013

New Sounds 2013
Montreal (Canada)
http://doe.concordia.ca/newsounds2013/

21–23 June, 2013

Approaches to Phonology and Phonetics (APAP)
Lublin (Poland)
http://apap.umcs.lublin.pl
apap2013@umcs.eu

25–26 June, 2013

Phonetics and Phonology in Iberia (PaPI 2013)
Lisbon (Portugal)
http://ww3.fl.ul.pt/laboratoriofonetica/papi2013/
papi2013@fl.ul.pt

07–10 July 2013

Phonetics and Phonology of Sub-Saharan Languages
Johannesburg (South Africa)
http://www.wits.ac.za/conferences/phonetics
Andrew.vanderspuy@wits.ac.za or Toni.borowsky@syndey.ac.au

22–27 July 2013

Word Stress: Dialectal Variation and Perception. Workshop of the International Congress of Linguists
Geneva (Switzerland)
http://www.cil19.org/en/workshops/word-stress-dialectal-variation-and-perception/
19icl@unige.ch

4–6 September, 2013

AEAL 2013 Bilbao, 7th International Conference on Language Acquisition
Bilbao (Spain)
http://www.aealbilbao.com/
info@aealbilbao.com

CALL FOR PAPERS

The *Phonetician* will publish peer-reviewed papers and short articles in all areas of speech science including articulatory and acoustic phonetics, speech production and perception, speech synthesis, speech technology, applied phonetics, psycholinguistics, sociophonetics, history of phonetics, etc. Contributions should primarily focus on experimental work but theoretical and methodological papers will also be considered. Papers should be original works that have not been published and are not being considered for publication elsewhere.

Authors should follow the *Journal of Phonetics* guidelines for the preparation of their manuscripts. Manuscripts will be reviewed anonymously by two experts in phonetics. The title page should include the authors' names and affiliations, address, e-mail, telephone, and fax numbers. Manuscripts should include an abstract of no more than 150 words and up to four keywords. The final version of the manuscript should be sent both in .doc and in .pdf files. It is the authors' responsibility to obtain written permission to reproduce copyright material.

All kinds of manuscripts should be sent in electronic form (.doc and .pdf) to the Editor. We encourage our colleagues to send manuscripts for our newly released section entitled MA research, which is a summary of the student's phonetics research describing their motivation, topic, goal, and results (no more than 1,200 words).

INSTRUCTIONS FOR BOOK REVIEWERS

Reviews in the *Phonetician* are dedicated to books related to phonetics and phonology. Usually the editor contacts prospective reviewers. Readers who wish to review a book mentioned in the list of "Publications Received" or any other book, should address the editor about it.

A review should begin with the author's surname and name, publication date, the book title and subtitle, publication place, publishers, ISBN numbers, price, page numbers, and other relevant information such as number of indexes, tables, or figures. The reviewer's name, surname, and address should follow "Reviewed by" in a new line.

The review should be factual and descriptive rather than interpretive, unless reviewers can relate a theory or other information to the book which could benefit our readers. Review length usually ranges between 700 and 2500 words. All reviews should be sent in electronic form to Prof. Judith Rosenhouse (e-mail: swantech@013.net).

ISPhS MEMBERSHIP APPLICATION FORM

Please mail the completed form to:

Treasurer:
Prof. Dr. Ruth Huntley Bahr, Ph.D.
Treasurer's Office:
Dept. of Communication Sciences and Disorders
4202 E. Fowler Ave. PCD 1017
University of South Florida
Tampa, FL 33620 USA

I wish to become a member of the International Society of Phonetic Sciences

Title: _____ Last Name: _____ First Name: _____
Company/Institution: _____
Full mailing address: _____
_____
Phone: _____ Fax: _____
E-mail: _____
Education degrees: _____
Area(s) of interest: _____

The Membership Fee Schedule (check one):

1. Members (Officers, Fellows, Regular)	$ 30.00 per year
2. Student Members	$ 10.000 per year
3. Emeritus Members	NO CHARGE
4. Affiliate (Corporate) Members	$ 60.000 per year
5. Libraries (plus overseas airmail postage)	$ 32.000 per year
6. Sustaining Members	$ 75.000 per year
7. Sponsors	$ 150.000 per year
8. Patrons	$ 300.000 per year
9. Institutional/Instructional Members	$ 750.000 per year

Go online at www.isphs.org and pay your dues via PayPal using your credit card.

☐ I have enclosed a cheque (in US $ only), made payable to ISPhS.

Date _____ Full Signature _____

Students should provide a copy of their student card

NEWS ON DUES

Your dues should be paid as soon as it convenient for you to do so. Please send them directly to the Treasurer:

Prof. Ruth Huntley Bahr, Ph.D.
Dept. of Communication Sciences & Disorders
4202 E. Fowler Ave., PCD 1017
University of South Florida
Tampa, FL 33620-8200 USA
Tel.: +1.813.974.3182, Fax: +1.813.974.0822
e-mail: rbahr@ usf.edu

VISA and MASTERCARD: You now have the option to pay your ISPhS membership dues by VISA or MASTERCARD using PayPal. Please visit our website, www.isphs.org, and click on the Membership tab and look under Dues for "paid online via PayPal." Click on this phrase and you will be directed to PayPal.

The Fee Schedule:

1. Members (Officers, Fellows, Regular)	$ 30.00 per year
2. Student Members	$ 10.00 per year
3. Emeritus Members	NO CHARGE
4. Affiliate (Corporate) Members	$ 60.00 per year
5. Libraries (plus overseas airmail postage)	$ 32.00 per year
6. Sustaining Members	$ 75.00 per year
7. Sponsors	$ 150.00 per year
8. Patrons	$ 300.00 per year
9. Institutional/Instructional Members	$ 750.00 per year

Special members (categories 6–9) will receive certificates; Patrons and Institutional members will receive plaques, and Affiliate members will be permitted to appoint/elect members to the Council of Representatives (two each national groups; one each for other organizations).

Libraries: Please encourage your library to subscribe to *The Phonetician*. Library subscriptions are quite modest – and they aid us in funding our mailings to phoneticians in Third World Countries.

Life members: Based on the request of several members, the Board of Directors has approved the following rates for **Life Membership** in ISPhS:

Age 60 or older:	$ 150.00
Age 50–60:	$ 250.00
Younger than 50 years:	$ 450.00