LUND UNIVERSITY

A brain network for integration of tone and suffix

Roll, Mikael; Söderström, Pelle; Horne, Merle

Published in:
[Publication information missing]

2015

Link to publication

Citation for published version (APA):
Roll, M., Söderström, P., & Horne, M. (2015). A brain network for integration of tone and suffix. *[Publication information missing]*, 140-140. http://www.uni-potsdam.de/morphproc2015/

Total number of authors:
3

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Papers 55. 2015 Linguistics Lund University

ISSN 0280-526X

Lund University
Centre for Languages and Literature

General Linguistics
Phonetics

Working Papers

55. 2015

Proceedings from *Fonetik 2015*
Lund, June 8–10, 2015

Lund University
Centre for Languages and Literature

General Linguistics
Phonetics

Working Papers

55. 2015

Proceedings from *Fonetik 2015*
Lund, June 8–10, 2015

Edited by Malin Svensson Lundmark, Gilbert Ambrazaitis and Joost van de Weijer

Working Papers
Department of Linguistics and Phonetics
Centre for Languages and Literature
Lund University
Box 201
S-221 00 LUND
Sweden
Fax +46 46 222 32 11
http://www.sol.lu.se/

This issue was edited by Malin Svensson Lundmark,
Gilbert Ambrazaitis and Joost van de Weijer

© 2015 The Authors and the Centre for Languages and Literature,
Lund University

ISSN 0280-526X

Printed in Sweden, Tryckeriet i E-huset, Lund, 2015

Preface

This volume of the Working Papers in General Linguistics and Phonetics contains the proceedings of the 28th Swedish Phonetics Conference, FONETIK 2015.

The conference is held from June 8 until June 10 2015 at Lund University in the recently inaugurated LUX building. LUX, located just beside SOL (the Centre for Languages and Literature), houses five departments within the humanities and theology. Together, SOL and LUX form a new campus for the humanities and theology at Lund University.

FONETIK 2015 is one in the series of annual conferences for phoneticians and speech scientists in Sweden, which regularly attract participants from Denmark, Finland and Norway and sometimes from other countries as well.

There are 20 contributions represented in this volume, many of which written by students and young academics. We interpret this as a sign that phonetic research in Sweden is very much alive, and therefore has a prosperous future.

This year's FONETIK is also an opportunity to remember Professor Gösta Bruce, a dear colleague, supervisor, mentor, and friend, who left us almost exactly five years ago. In 2010, Gösta was part of the organizing committee for the Lund edition of the FONETIK conference, but he was then unable to attend the meeting because of his illness. Gösta always highly appreciated the FONETIK conferences and, according to oral tradition, he was the only phonetician in Sweden who attended every single meeting, from their beginning in 1986 until 2009. We believe that Gösta would have loved to witness the new generation of phoneticians in Sweden.

We hope that this year's LundaFONETIK, organized for the first time without the help of Gösta Bruce, will be as successful as usual! We thank all contributors to the proceedings, and we acknowledge the valuable support from The Swedish Phonetics Foundation (Fonetikstiftelsen), IDO foundation (Stiftelsen Lundbergska Idofonden), SOL (the Centre for Languages and Literature) and LUX.

Lund, June 2015

The Organizing Committee

Gilbert Ambrazaitis, Johan Frid, Anastasia Karlsson, Malin Svensson Lundmark, Susanne Schötz, Mechtild Tronnier and Joost van de Weijer

Previous Swedish Phonetics Conferences (from 1986)

I	1986	Uppsala University
II	1988	Lund University
III	1989	KTH Stockholm
IV	1990	Umeå University (Lövånger)
V	1991	Stockholm University
VI	1992	Chalmers and Gothenburg University
VII	1993	Uppsala University
VIII	1994	Lund University (Höör)
—	1995	(XIIIth ICPhS in Stockholm)
IX	1996	KTH Stockholm (Nässlingen)
X	1997	Umeå University
XI	1998	Stockholm University
XII	1999	Gothenburg University
XIII	2000	Skövde University College
XIV	2001	Lund University (Örenäs)
XV	2002	KTH Stockholm
XVI	2003	Umeå University (Lövånger)
XVII	2004	Stockholm University
XVIII	2005	Gothenburg University
XIX	2006	Lund University
XX	2007	KTH Stockholm
XXI	2008	Gothenburg University
XXII	2009	Stockholm University
XXIII	2010	Lund University
XXIV	2011	KTH Stockholm
XXV	2012	Gothenburg University
XXVI	2013	Linköping University
XXVII	2014	Stockholm University

Contents

Inhalation amplitude and turn-taking in spontaneous Estonian conversations **1**
Kätlin Aare, Marcin Włodarczak & Mattias Heldner

What affects recognition most – wrong word stress or wrong word accent? **7**
Åsa Abelin & Bosse Thorén

Multimodal levels of prominence: a preliminary analysis of head and eyebrow movements in Swedish news broadcasts **11**
Gilbert Ambrazaitis, Malin Svensson Lundmark & David House

Transfer of rising tone patterns from Swedish L1 into Spanish L2 – which are the communicative consequences? **17**
Berit Aronsson

The realisation of sj- and tj-sounds in Estonian Swedish: some preliminary results **23**
Eva Liina Asu, Otto Ewald & Susanne Schötz

Grimaldi's "Discovery of the Cat Language": A theory in need of revival (or perhaps not?) **27**
Robert Eklund

Languages with pulmonic ingressive speech: updating and adding to the list **31**
Robert Eklund

A spectral analysis of the backing of Afrikaans /s/ in the consonant cluster /rs/ **35**
Otto Ewald

Self-perception of vocal and articulatory effort in consonant production by native Swedish speakers **41**
Iris Gordon-Bouvier, Josefine Kyhle, Anita McAllister, Hanna Norman, Sarah Paues, Camille Robieux & Sofia Strömbergsson

Temporal aspects of breathing and turn-taking in Swedish multiparty conversations **47**
Jonna Hammarsten, Roxanne Harris, Nilla Henriksson, Isabelle Pano, Mattias Heldner & Marcin Włodarczak

Deaccented verbs in Swedish **51**
Anna Hed & Anna Smålander

A change in the openness of two vowel phoneme pairs in eastern Icelandic: An acoustic analysis of /œ:/ vs. /ʏ:/, and /ɛ:/ vs. /ɪ:/ **57**
Guðlaug Hilmarsdóttir

On the temporal domain of co-speech gestures: syllable, phrase or talk spurt? **63**
David House, Simon Alexanderson & Jonas Beskow

Intonations and functions of questions in Helsinki Swedish conversations **69**
Martina Huhtamäki

A pilot study: acoustic and articulatory data on tonal alignment in Swedish word accents **75**
Malin Svensson Lundmark, Johan Frid & Susanne Schötz

An acoustic analysis of the cattle call "kulning", performed outdoors at Säter, Dalarna, Sweden **81**
Anita McAllister & Robert Eklund

Agonistic vocalisations in domestic cats: a case study **85**
Susanne Schötz

Using tonal cues to predict inflections **91**
Pelle Söderström, Merle Horne & Mikael Roll

Foreign accent: influences of the sound system of Serbian on the production of Swedish L2 **95**
Mechtild Tronnier & Elisabeth Zetterholm

Halfway to Estuary English with H. G. Wells (1866-1946) **101**
Sydney A.J. Wood

Inhalation amplitude and turn-taking in spontaneous Estonian conversations

Kätlin Aare, Marcin Włodarczak and Mattias Heldner
Department of Linguistics, Stockholm University, Stockholm

Abstract

This study explores the relationship between inhalation amplitude and turn management in four approximately 20 minute long spontaneous multiparty conversations in Estonian. The main focus of interest is whether inhalation amplitude is greater before turn onset than in the following inhalations within the same speaking turn.

The results show that inhalations directly before turn onset are greater in amplitude than those later in the turn. The difference seems to be realized by ending the inhalation at a greater lung volume value, whereas the initial lung volume before inhalation onset remains roughly the same across a single turn. The findings suggest that the increased inhalation amplitude could function as a cue for claiming the conversational floor.

Introduction

Previous research has shown that speech planning is reflected in respiratory patterns, at least in read speech. For instance, the duration and amplitude of inhalation have been found to correlate positively with the upcoming utterance length in several studies (e.g. Winkworth, Davis, Adams, Ellis, 1995; Fuchs, Petrone, Kirvokapić, Hoole, 2013). Also the location of inhalation is strongly determined by speech planning. Almost all inhalations in read speech occur at major constituent boundaries, such as paragraphs, sentences or phrases (Conrad, Thalacker, Schönle, 1983; Grosjean and Collins, 1979).

By contrast, breathing in spontaneous speech shows a less consistent pattern. It has been claimed that as many as 13% of all inhalations in spontaneous monologues occur at grammatically inappropriate locations (Wang, Green, Nip, Kent, Kent, 2010), possibly due to the additional demands of real-time speech planning. The effect should be even more pronounced in spontaneous conversation, where the communicative demands are different.

A key characteristic of the conversational rhythm is its oscillating pattern – generally, one speaker at a time has the speaking turn and longer stretches of simultaneous speech tend to be avoided. Thus, the exchange of speaker and listener roles needs to be precisely coordinated by means of turn-taking cues indicating the intention to take, hold or release the turn (McFarland, 2001). Breathing patterns have previously been hypothesized to be part of the turn-taking system. Inhalations have been claimed to be an interactionally salient cue to speech initiation (Schegloff, 1996) and to be deeper before turn initiation (Ishii, Otsuka, Kumano, Yamato, 2014). Finally, breath holding and exhalation have been suggested as turn keeping and turn-yielding devices, respectively (French and Local, 1983).

Furthermore, durational properties of respiration have been shown to reflect turn-taking intentions. Speakers tend to minimise pause durations inside the turn by inhaling more quickly, and by reducing the delay between inhalation offset and speech onset (Rochet-Capellan and Fuchs, 2014; Hammarsten, Harris, Henriksson, Pano, Heldner, Włodarczak, this volume). As inhalation duration and depth have been found to correlate, especially in read speech (e.g. Rochet-Capellan and Fuchs, 2013), the amplitude of non-initial inhalations in a speaking turn should also be smaller.

Therefore, in addition to being governed by the demands of phrasing and speech planning, breathing patterns in spontaneous conversation may depend, at least partly, on speaker's communicative goals linked to claiming and keeping the turn. In this study, we investigate whether turns consisting of several breathing cycles show a pattern where the turn-initiating inhalation amplitude is greater than the amplitude of the following inhalations within the turn.

Method

Data acquisition was carried out by recording respiratory activity synchronized with audio and video in spontaneous three-party conversations each lasting approximately 22 minutes. Participants were 20- to 35-year-old (with the mean of 26) healthy native speakers of Estonian with an average Body Mass Index of 21.9. The speakers did not report any history of speech, language, hearing or respiratory disorders, and had never been smokers. They were invited to travel to Stockholm and take part in the recording sessions by personal communication. The speakers in each recording session had known each other for a period from a few weeks to 14 years. With the exception of two couples living together, the speakers described their relationships as friends.

The recordings took place in the Phonetics Laboratory at Stockholm University. The participants had no knowledge of the exact aim of the experiment prior to the recording and they were free to choose the topics of conversations. They were instructed to wear tight-fitting clothes to minimise distortions in the respiratory signals.

Respiratory activity was measured with Respiratory Inductance Plethysmography (Watson, 1980), which quantifies changes in the rib cage and abdomen cross-sectional areas by means of two elastic transducer belts (Ambu RIP-mate) placed at the level of the armpits and the navel. The overall lung volume change was estimated by isovolume manoeuvre (Konno and Mead, 1967). A more detailed setup description is provided in Edlund, Heldner and Włodarczak (2014).

The respiratory signal was recorded using PowerLab (ADInstruments, 2014). Audio was captured using head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). The speakers were asked to stand around a 1-meter-high table, each facing a GoPRO Hero 3+ camera recording the upper part of the torso, and to avoid large movements.

Annotation of the data was carried out semi-automatically using Praat (Boersma and Weenink, 2015) and Python scripts (Buschmeier and Włodarczak, 2013). The sum of the rib cage and abdomen signals was used to segment the breathing signal into periods of inhalations and exhalations (for details see: Włodarczak and Heldner, in press). A total of 11.9% of the automatically assigned borders were either moved or added manually due to inaccuracy of the automatic annotation. In addition, following Jaffe and Feldstein (1970), silences and overlaps were classified depending on whether they coincided with speaker change or were followed by more speech from the same speaker. Accordingly, the speaking turn has been defined as an uninterrupted series of speech segments from a single speaker. As backchannels, coughing and laughter are commonly considered to be non-interruptive, they were not classified as claiming a turn. This is particularly true of backchannels, which are unplanned and produced by the listener to give short feedback to the speaker (see e.g. Heldner, Hjalmarsson, Edlund, 2013; Aare, Włodarczak, Heldner, 2014). Consequently, only uninterrupted speaking turns that included multiple breath groups were analysed. The amount of data was limited further by excluding some speech stretches that coincided partly with inhalations.

Lung volume levels were normalised with respect to speaker's minimum and maximum lung volumes measured at the calibration stage of the recording. These values correspond to vital capacity (VC), the maximum volume of air exhaled after a maximum inhalation (Hixon, 2006). The final part of analysis was carried out with R (R Core Team, 2015).

Results

Data distribution

Due to the filtration procedure, 50 suitable speaking turns were left for analysis. These turns consisted of 128 breath groups produced by 11 speakers. Table 1 illustrates the distribution of the number of breath groups and inhalations for one speech turn. All multi-breath-group turns included at least two inhalation phases but rarely more. Therefore, to ensure sufficient sample sizes, the results below are based on the data from the first two inhalations in each of the 50 turns.

Table 1. Inhalations per speaking turn.

Number of inhalations	Frequency
2	31
3	10
≥ 4	9
Σ	50

Inhalation start and end levels

The initial and final lung volume levels for the first two inhalations in a turn, normalised to speaker's vital capacity (%VC), are shown in Figure 1.

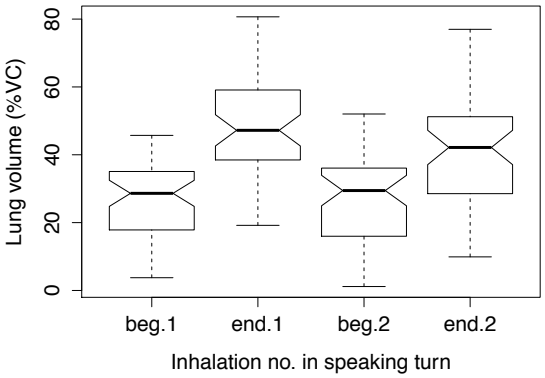

Figure 1. A comparison of the start and end lung volume levels for the first two inhalations in a turn. All data is normalised to speaker's vital capacity. "1" marks the first inhalation of a turn, occurring before turn onset, and "2" marks the second inhalation, located inside the speaking turn; "beg" and "end" indicate the point of measurement – the beginning or end of the corresponding inhalation.

Figure 1 shows that while the initial lung volumes in the first and second inhalation are practically identical (except for a slightly larger range in the latter), the end values differ more considerably. Specifically, the second inhalation's end value is lower in median and both range end points, and is characterised by larger range. Furthermore, the difference between the first and second inhalation's end lung volume means is statistically significant *($t(98) = 2.262$, $p = .026$)*.

A more detailed relationship of the start and end lung volume levels for both inhalations can be seen on Figure 2. The figure shows a marked positive relationship (correlation: *$r(48) = .657$, $p = .000$*) between the inhalation start and end lung volume levels for both inhalations.

Inhalation amplitude

Inhalation amplitudes (also normalised with respect to speakers' vital capacity) for turn-initial and turn-medial inhalations are presented in Figure 3. The overall inhalation amplitude falls into the range between 26 and 45% of VC.

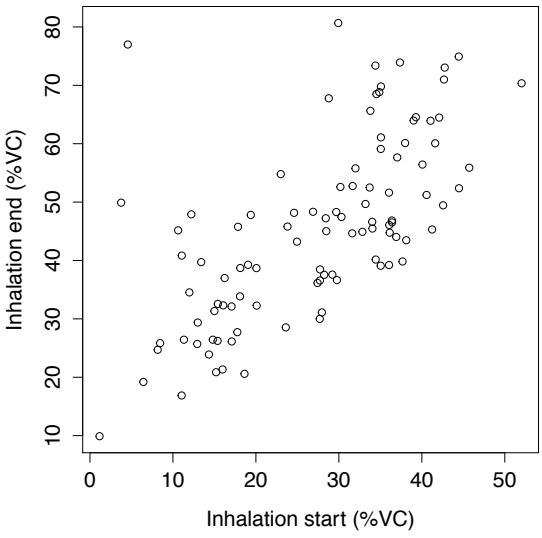

Figure 2. A scatterplot of the normalized inhalation start and end lung volume levels (%VC).

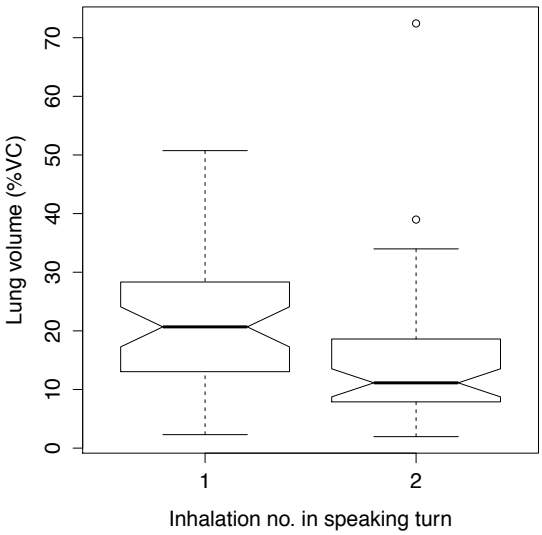

Figure 3. Inhalation amplitude for the first and second inhalation in a turn.

As can be seen, the first and second inhalation in a turn have different amplitude medians and ranges. The turn-initial inhalation exhibits higher and more symmetrically distributed amplitude values with values ranging to around 50% of vital capacity and a median of 20% VC. The following inhalation has a considerably lower median and a smaller range. The difference between mean amplitude values is also significant *($t(98) = 2.854$, $p = .005$)*.

Discussion

The comparison of turn initial and turn medial inhalations in a single turn shows that the turn-initial inhalation is significantly greater in amplitude. As can be seen in Figure 1, speakers tend to initiate inhalations at approximately the same level in the vital capacity, but inhale more deeply before taking the turn.

This might imply that a deeper inhalation cues the intention to take the turn. As the larger amplitude is accomplished by reaching a higher lung volume, the cue could just as well be the end lung volume level, rather than the amplitude itself. In other words, the actual change in amplitude may be less important than the end level within the vital capacity. We leave this hypothesis for future research.

We may also speculate that a deeper inhalation projects a longer turn, possibly spanning several breath cycles. Simply put, by signalling an intention to produce a longer contribution, speakers might minimise competition for the conversational floor at turn internal pauses. At the same time, more shallow inhalations at turn medial pauses could serve as a phrasing device to indicate how the speaker intended to structure the message, thereby facilitating listener's understanding.

There are a number of additional factors that could provide further insight into the interplay between turn-taking and respiration. For instance, shorter inhalations might require higher airflow, which in turn increases the likelihood of fricative noise. This audible inhalation in itself might also function as a turn-holding or phrasing device. We plan to address this issue in subsequent studies.

Conclusions

This exploratory study focused on the physical constraints governing speech breathing in connection with turn-taking in spontaneous multiparty conversations among native speakers of Estonian. The examination of lung volumes in speech turns spanning several breath cycles indicates that turn initial inhalations are deeper than inhalations later in the turn. This is accomplished by inhaling to a higher lung volume during the first inhalation, with the pre-inhalatory lung volumes remaining relatively stable across consecutive inbreaths.

The use of respiratory cues for the organization of turn-taking is not a new discovery, but this work contributes novel findings regarding lung volumes. Although the limited size of the data set restricts the possible conclusions, the patterns discovered in this study indicate that inhalation depth is sensitive to speakers' intention to start or continue speaking. The results thus provide another evidence in favour of breathing serving as a potentially important turn-taking cue in spontaneous conversation.

Acknowledgements

The research presented here was funded in part by the Swedish Research Council project 2014-1072 *Andning i samtal* (*Breathing in conversation*).

References

Aare K., Włodarczak M. and Heldner M. (2014). Backchannels and breathing. In: M. Heldner (Ed.), *Proceedings of FONETIK 2014.* Stockholm, Sweden, 47-52.

ADInstruments. (2014). LabChart software and PowerLab hardware (Version 8). New South Wales, Australia: ADInstruments

Boersma P. and Weenink D. (2015). Praat: doing phonetics by computer [Computer program] (Version 5.3.84). Retrieved from http://www.praat.org/

Buschmeier H. and Włodarczak M. (2013). TextGridTools: A TextGrid Processing and Analysis Toolkit for Python. In P. Wagner (Ed.), *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013),* 152-157. Dresden: TUDpress.

Conrad B., Thalacker S. and Schönle P. (1983). Speech respiration as an indicator of integrative contextual processing. *Folia Phoniatrica et Logopaedica* 35, 220-225.

Edlund J., Heldner M. and Włodarczak M. (2014) Catching wind of multiparty conversation. In: J. Edlund, D. Heylen and P. Paggio (Eds.), *Proceedings of Multimodal Corpora: Combining applied and basic research targets (MMC 2014)*. Reykjavík, Iceland.

French P. and Local J. (1983). Turn-competitive incomings. *Journal of Pragmatics* 7, 17-38.

Fuchs S., Petrone C., Krivokapić J. and Hoole P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, 29-47.

Grosjean F. and Collins M. (1979). Breathing, pausing and reading. *Phonetica* 36(2), 98-114.

Hammarsten J., Harris R., Henriksson N., Pano I., Heldner M. and Włodarczak M. (this volume). Temporal aspects of breathing and turn-taking in Swedish multiparty conversations.

Heldner M., Hjalmarsson A. and Edlund J. (2013). Backchannel relevance spaces. In E.L. Asu and P. Lippus (Eds.), *Nordic Prosody: Proceedings of the*

XIth Conference, Tartu 2012, 137-146. Frankfurt am Main, Germany: Peter Lang.

Hixon T.J. (2006). *Respiratory function in singing: A primer for singers and singing teachers*. Tucson, Arizona: Redington Brown.

Ishii R., Otsuka K., Kumano S. and Yamato J. (2014). Analysis of respiration for prediction of "who will be next speaker and when?" in multi-party meetings. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14*), 18-25.

Jaffe J. and Feldstein S. (1970). *Rhythms of dialogue*. New York, NY, USA: Academic Press.

Konno K. and Mead J. (1967). Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology*, 22(3), 407–422.

McFarland D.H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* 44, 128-143.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org/.

Rochet-Capellan A. and Fuchs S. (2013). The interplay of linguistic structure and breathing in German spontaneous speech. In *Proceedings of Interspeech*, 1128–1132.

Rochet-Capellan A. and Fuchs S. (2014). Take a breath and take the turn: how breathing meets turns in spontaneous dialogue. In R. Smith, T. Rathcke, F. Cummins, K. Overy and S. Scott (Eds.), *Philosophical Transactions of the Royal Society B: Biological Sciences,* 369(1658), 1-10.

Schegloff E.A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E.A. Schegloff and S.A. Thompson (Eds.), *Interaction and Grammar*, 52-133. Cambridge: Cambridge University Press.

Watson H. (1980). The technology of respiratory inductive plethysmography. In F.D. Stott, E.B. Raftery and L. Goulding (Eds.), *Proceeding of the Second International Symposium on Ambulatory Monitoring (ISAM 1979)*. London: Academic Press.

Winkworth A.L., Davis P.J., Adams R.D. and Ellis E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech & Hearing Research* 38, 124-144.

Wang Y.T., Green J.R., Nip I.S., Kent R.D. and Kent J.F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. *Folia Phoniatrica et Logopaedica* 62 (6), 297-302.

Włodarczak M. and Heldner M. (in press). Respiratory properties of backchannels in spontaneous multiparty conversation. In *Proceedings of ICPhS 2015.*

What affects recognition most – wrong word stress or wrong word accent?

Åsa Abelin[1] and Bosse Thorén[2]
[1]Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg
[2]School of Humanities and Media Studies, Dalarna University

Abstract

In an attempt to find out which of the two Swedish prosodic contrasts of 1) word stress pattern and 2) tonal word accent category has the greatest communicative weight, a lexical decision experiment was conducted: in one part word stress pattern was changed from trochaic to iambic, and in the other part trochaic accent II words were changed to accent I.

Native Swedish listeners were asked to decide whether the distorted words were real words or 'non-words'. A clear tendency is that listeners preferred to give more 'non-word' responses when the stress pattern was shifted, compared to when word accent category was shifted. This could have implications for priority of phonological features when teaching Swedish as a second language.

Introduction

This study started with a discussion at the Phonetics meeting 2014. The topic concerned Swedish spoken with a foreign accent, and whether wrong word stress or wrong word accent was most detrimental for recognition and understanding of words. When teachers make curricula for second language speakers, they are helped by knowing which phonetic or phonological features are more or less crucial for the understanding of speech. There are some structural and anecdotal evidence that word stress should play a more important role in the perception and understanding of Swedish than the tonal word accent. The aim of the study is to find out which of two distortions causes the most difficulty in identifying some disyllabic words: 1) changing the word stress category from trochaic to iambic or 2) changing the tonal word accent category from accent II to accent I.

Background

Swedish word stress is about prominence contrasts between syllables, mainly signalled by syllable duration (Fant & Kruckenberg 1994), although F0 gestures, voice source parameters and differences in vowel quality combine to signal syllable prominence (ibid.) The tonal word accent, however, is mainly signalled by changes in the F0 curve and the timing of those changes within the word. According to Bruce (1977, 2012) and Elert (1970), word stress in Swedish is variable, and words can have different meanings depending on where the main stress is placed, as found in *ˈbanan* 'the path/course' and *baˈnan* 'banana'. A great number of disyllabic trochaic-iambic minimal pairs can be created. A smaller number of trisyllabic minimal pairs, such as *ˈIsrael* 'the state of Israel' and *israˈel* 'Israeli citizen', are also possible.

According to standard accounts Swedish has two word accent categories, accent I (acute), e.g. *tómten* 'the plot', and accent II (grave), *`tomten* 'Santa Claus', cf. Elert (1970), even though only the grave accent can be considered a real word accent. It is the only one of these two that predicts that the main stressed syllable and the following syllable belong to the same word (in a two-syllable word) i.e. having a cohesive function, and it is limited to the word, simple or compound. The word accent is connected with a primary stressed syllable. In isolation the words usually carry sentence accent and accent II then tends to involve two F0 peaks.

Method

Material and design

The material consisted of 10 trochaic (accent I) words, e.g. *bílen* 'the car', 10 originally trochaic words pronounced with iambic stress, e.g. *vägén* 'the road', 10 iambic words, e.g. *kalás* 'the party', 10 accent II words, e.g. *`gatan* 'the street', 10 originally Accent II words pronounced with trochaic stress accent I, e.g. *´sagan* 'the fairy

tale', and finally 26 disyllabic non-words, with varying stress or tonal accents. All the words were nouns in the definite form (with one exception) apart from the iambic words. The words were recorded by a male phonetician with a neutral dialect. Recording and editing was made with the software Praat (Boersma & Weenink, 2013).

There was some deliberation about how to treat vowel quality in the stressed and unstressed syllables, since these vary according to degree of stress. We decided to choose vowels which do not vary so much in unstressed vs. stressed position, e.g. /e/ rather than /a/, and keep the quality of the original word, e.g. not changing [e] to [ɛ] in unstressed position. Each word was presented until it self-terminated, in all cases below 1000 ms. Simultaneously the subjects had 1000 ms to react to each stimulus. The time allotted for reaction to the stimuli thus started when the word started. Between each word there was a 1000 ms pause.

For building and running the experiment the software PsyScope was used (Cohen, MacWhinney, Flatt & Provost, 1993).

Procedure

A lexical decision test was performed where 18 female L1 speakers of Swedish, approximately 20–25 years of age, heard the above described 76 words, one by one in random order. The subjects were instructed to press one key on a keyboard if the word was a real word and another key if the word was a non-word. The subjects were instructed to decide as quickly as possible, whether the word they heard was a real word or not. Reactions that were not registered within the 1000 ms period were categorized as loss.

Results

Accuracy

Figure 1 shows the main results of the experiment. It turned out that the task was quite difficult, and that the loss in the experiment was large.

Loss signifies that the reaction times were too long, over 1000 ms. The difficulty of responding quickly could be due to the fact that the word stimuli were quite long, between 660 and 979 ms. However, the main result does not concern the reaction times, but the difference in assessment of word status of the words with wrongly pronounced tonal word accent, compared with the words with wrongly pronounced stress placement.

Figure 1 shows that the mispronounced words that were generally judged as real words were the accent II words pronounced with accent I (23% 'yes' responses and 3% 'no' responses), while the words that were generally judged as non-words were the trochaic accent I words pronounced as iambic accent I (7% 'yes' responses and 19% 'no' responses). This suggests that identification and comprehensibility of speech is more affected by wrong stress

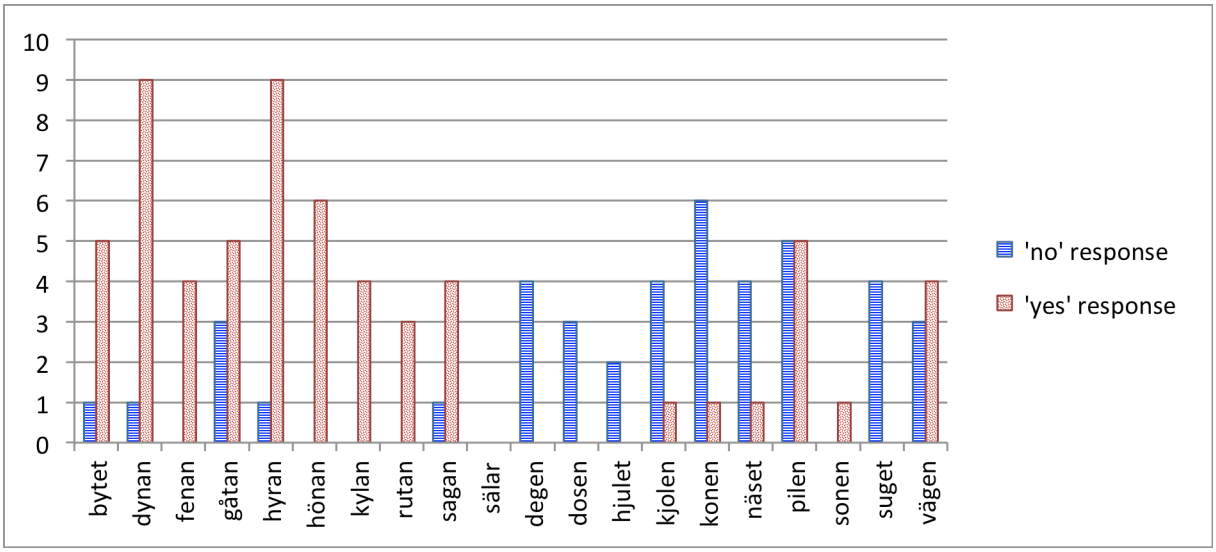

Figure 1. Number of persons deciding on mispronounced words being real words (dotted bars) or non-words (striped bars). The first 10 words to the left are accent II words pronounced with accent I, and the next 10 words, to the right, are trochaic accent I words pronounced with iambic accent.

placement in comparison with wrong word accent. Furthermore, there was a larger loss for the words with wrong stress placement than for the words with wrong tonal word accent, however not significant.

Figure 2 compares the wrongly pronounced words with the correctly pronounced words. The figure shows that the correctly pronounced words are the most robust; they exhibit a smaller loss and they are more often assessed as real words. The words which were most frequently misjudged were the words with wrong stress placement.

There is interaction between loss, 'no' responses and 'yes' response: Where there are more 'no' responses the loss is greater. This could be due to the simple fact that 'no' responses generally have longer reaction times than 'yes' responses; thus, it could be that in some cases when a 'no' response is intended the response time exceeds 1000 ms. But the result could also be due to an impossibility to interpret the wrongly pronounced word.

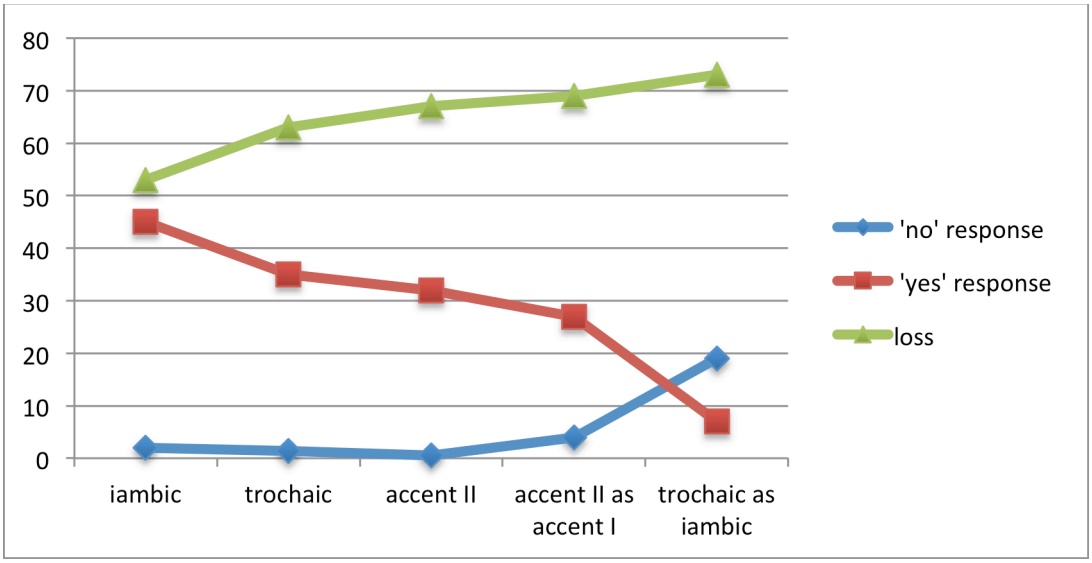

Figure 2. Mean percent for decisions on mispronounced words being real words (squares) or non-words (diamonds), in comparison with correctly pronounced words; iambic, trochaic and accent 2. Triangles stand for loss.

Reaction times

There was not a large difference in mean reaction time between the two wrongly pronounced groups. Mean value for mispronounced accent II was 877 ms and mean value of mispronounced trochaic accent I was 915 m sec. This is not a significant difference, but it is nevertheless like comparing apples with pears, i.e. actually comparing 'yes' responses with 'no' responses, which are generally longer. To compare reaction times for the 'yes' responses is not relevant since there were so few 'yes' responses for the words with wrong stress placement.

Durations of sound stimuli

The durations of the sound stimuli were measured, and we found that the wrongly pronounced trochaic accent I words, pronounced as iambic, were slightly longer. However this did not correlate with reaction times.

In general, reaction times were longer than the word durations, but not if deducting 200 ms for motor action. There is a tendency that when the durations are shorter, loss is smaller and the 'yes' responses are more numerous.

Discussion and conclusion

The results can be discussed in relation to "left-to-right" models of speech perception and to where the actual recognition point is situated (cf. Marslen-Wilson, 1987). Is prosody a factor here alongside segmental information? One question is whether an early absence of stress placement would be more detrimental for recognition than a late absence, i.e. would a stress-placement-changed trochaic word (which ought to have stress on the first syllable) be more difficult to process than a stress-placement-changed iambic word (which ought to have stress on the second

syllable)? And similarly, would accent I words pronounced with accent II be more difficult to recognize than accent II words pronounced with accent I? Such experiments are presently carried out and analyzed. Preliminary results show the same general tendency as in the present experiment; misplaced stress is more detrimental to recognition and yields more 'no' responses than words pronounced with the wrong tonal word accent. There is also a tendency for longer reaction times for misplaced stress compared with mispronounced word accent.

The words of the present experiment were not checked for frequency or number of phonological neighbours. It could be the case that some of the iambic words (which often are loan words) have a lower frequency. On the other hand, the correctly pronounced iambic words were the words that had the least loss, the highest number of 'yes' responses and the lowest number of 'no' responses, which might indicate an effect of few phonological neighbours, as concerns "stress related neighbours". The reason that words were not balanced for frequency was that it was difficult to find suitable words. However, frequency is not a main issue since the results mainly concern correct interpretation or misinterpretation, not reaction time.

Another reflection is the following: What does it entail that the iambic (correct) words are not in the definite form? Morphology, such as different inflectional forms, can affect processing. Söderström (2012) studied perception of accent I and accent II in a mismatch condition where accent I words were followed by accent II inducing suffixes, and accent II words were followed by accent I inducing suffixes. He found that there is a stronger relation between suffixes and accent II compared with accent I, which could imply that accent II could indeed be very important to perception, identification and comprehension in certain contexts.

In relation to the studies of Söderström (2012), Söderström, Roll & Horne (2012) the question arises whether accent II might be more important to comprehension where there are other errors, e.g. in the speech of learners of Swedish as a second language, which might use the wrong suffixes on nouns or verbs. Adding further learner errors such as word order mistakes or wrong lexical choices the picture becomes complicated.

We are well aware of that our experiment does not show high ecological validity since it tested deliberately mispronounced words which were judged out of context. Follow-up studies will hopefully be made in more natural scenarios.

However, the present results suggest that learners of Swedish as a second language benefit more from proficiency in stress placement than in choice of word accent category or precise realization of word accent category.

This is also indicated by the fact that word accent categories are realized differently in different geographical regions, and that some varieties do not utilize the contrast at all.

References

Boersma, P. & Weenink, D. (2013). Praat: Doing phonetics by computer (http://www.praat.org).

Bruce, G. (1977). *Swedish word accents in sentence perspective* (Vol. 12). Lund University.

Bruce, G. (2012). *Allmän och svensk prosodi*, Studentlitteratur, Lund.

Cohen J.D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257-271.

Elert, C.-C. (1970). *Ljud och ord i svenskan*, Stockholm: Almqvist & Wiksell.

Fant, G. & Kruckenberg, A. (1994). Notes on stress and word accent in Swedish *STLQPSR*. 2-3.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25: 71–102.

Söderström, P. (2012). *Processing Swedish word accents - evidence from response and reaction times*, MA thesis in General linguistics, Lund University.

Söderström, P., Roll, M., & Horne, M., (2012). Processing morphologically conditioned word accents. *Mental Lexicon* 7, 77–89.

Thorén, B. (2008). *The priority of temporal aspects in L2-Swedish prosody: Studies in perception and production*. PhD thesis, Stockholm University.

Multimodal levels of prominence: a preliminary analysis of head and eyebrow movements in Swedish news broadcasts

Gilbert Ambrazaitis[1], Malin Svensson Lundmark[1] and David House[2]
[1]Centre for Languages and Literature, Lund University
[2]Department of Speech, Music and Hearing, KTH Stockholm

Abstract

This paper presents a first analysis of the distribution of head and eyebrow movements as a function of (a) phonological prominence levels (focal, non-focal) and (b) word accent (Accent 1, Accent 2) in Swedish news broadcasts. Our corpus consists of 31 brief news readings, comprising speech from four speakers and 986 words in total. A head movement was annotated for 229 (23.2%) of the words, while eyebrow movements occurred much more sparsely (67 cases or 6.8%). Results of χ^2-tests revealed a dependency of the distribution of movements on the one hand and focal accents on the other, while no systematic effect of the word accent type was found. However, there was an effect of the word accent type on the annotation of 'double' head movements. These occurred very sparsely, and predominantly in connection with focally accented compounds (Accent 2), which are characterized by two lexical stresses. Overall, our results suggests that head beats might have a closer association with phonological prosodic structure, while eyebrow movements might be more restricted to higher-level prominence and information-structure coding. Hence, head and eyebrow movements can represent two quite different modalities of prominence cuing, both from a formal and functional point of view, rather than just being cumulative prominence markers.

Introduction

People gesture while they speak, using various parts of the body. Hands, the head and certain facial areas (such as the eyebrows) have so far received most attention in research on multimodal communication, i.e. the interaction of gestures and speech. While a single gesture often serves several functions at once, a basic typology of gestures would include emblems, iconic gestures, deictic gestures and beat gestures (Casasanto, 2013). In this typology, beat gestures are special in that they do not necessarily convey any semantic content. Beat gestures are generally understood as simple, rapid movements of a hand, a finger, the head, or the eyebrows, "often repeated, and timed with prosodic peaks in speech" (*Ibid*., p. 373). Moreover, "the cognitive and communicative functions of beats are not well understood" (*Ibid*., p. 373).

However, recent studies have begun solving a couple issues concerning beat gestures: For instance, it has been shown that beat gestures can facilitate both speech production (Lucero et al., 2014) and speech processing (Biau and Soto-Faraco, 2013; Wang and Chu, 2013).

A growing body of evidence also suggests that hand, head and eyebrow movements are aligned with pitch accents in speech and in this way contribute to the production and perception of prosodic prominence (Yasinnik et al., 2004; Flecha-García, 2010). For instance, in a database of video recordings of two shorter (5-7.5 minute) academic lectures by two male speakers of English, Yasinnik et al. (2004) found that beat gestures by the hands, head, or eyebrows occurred in close alignment with ToBI-labelled pitch accents in about 65-90% of the cases. Similarly, Flecha-García (2010) found that eyebrow raises preceded pitch accents by on average 60 ms in a corpus of English face-to-face dialogue.

This study is part of the research project *Multimodal levels of prominence*, investigating the interaction of head movements, eyebrow movements and pitch accents at the sentence level (so-called *focal accents* in the Swedish tradition), in Stockholm Swedish. The main research question of the project is whether the

three modalities (pitch accent, head, eyebrows) can interact in various ways in order to produce different levels of prominence, and whether these prominence levels are used to encode different shades of information structure (such as *new* vs. *accessible* information).

This paper presents a first analysis of the distribution of head and eyebrow movements as a function of verbal prominence levels and word accent categories in a corpus of news readings from Swedish Television. Our approach is inspired by Swerts and Krahmer's (2010) study on Dutch newsreaders, which argues that news readings "represent natural data that are still sufficiently constrained to be able to explore specific functions of their expressive style" (p. 198). Swerts and Krahmer (2010) found that the more accented a word was on an auditory scale (no accent, weak accent, strong accent), the more likely the word was to also be accompanied by a head movement, an eyebrow movement or both (most common in the strongly-accented words). While our materials are largely comparable to theirs, the studies differ in the important respect that we – in this first step – did not establish a perceptual prominence rating. Instead, we are making use of the fact that Swedish has two phonological prosodic prominence levels, which can rather easily be distinguished when inspecting the fundamental frequency contour. Thus, our point of departure is the question whether *phonological* prominence levels – which often, but not always necessarily reflect perceptual prominence levels – have an effect on the distribution of head and eyebrow movements.

Unlike so-called intonation languages like English and German, Swedish is a pitch-accent language, making use of pitch contrasts at the lexical level. In particular, Swedish has a binary distinction between two word accents (Accent 1 and Accent 2), two different pitch accents assigned to words by means of lexical/ morphological rules. In addition, words can be highlighted at the sentence level, just as in English or German. For Stockholm Swedish, a phonological distinction is generally assumed between the non-focal, accented realization of a word (tonal pattern in Stockholm Swedish: H[igh]-L[ow]; with a different timing of the HL for Accent 1 and 2, cf. Bruce, 1977), and a focal realization of a word (HLH, i.e. an additional High tone). Note: While the non-focal vs. focal accents represent two different phonological prominence levels, no difference in prominence is generally assumed between the two word accents (Accent 1 vs. 2).

Therefore, the hypothesis was that focally accented words would coincide with head or eyebrow movements more often than non-focal words, while the word accent category (Accent 1 vs. 2) should have no effect of the distribution of head or eyebrow movements.

Method

Audio and video data of 31 brief news readings from Swedish Television (SVT Rapport, 2013) were analyzed. The corpus included speech from four newsreaders: two female (Sofia Lindahl, Katarina Sandström) and two male (Pelle Edin, Alexander Norén). Each piece of news typically contained 1-3 sentences, amounting to 986 words in total. The recordings were retrieved on DVD from the National Library of Sweden (Kungliga Biblioteket).

The material was transcribed, segmented at the word level, and annotated using ELAN (Wittenburg et al., 2006). Word segmentations were adjusted using Praat (Boersma and Weenink, 2014) and re-imported in ELAN prior to doing the annotations. Head and eyebrow movements, as well as focal accents were labelled manually by three annotators. In the analysis, a word was counted as coinciding with one of the three events (focal accents, head, eyebrow movements) in the event of an agreement between at least two annotators. The annotation scheme was simple in that only the presence vs. absence of any of the three events was judged upon. That is, no time-aligned annotations were made for the purpose of this study, and hence, no decisions had to be made upon temporal onsets and offsets of the movements. A word was annotated for bearing a (head or eye-brow) movement in the event that the head or at least one eyebrow rapidly changed its position, roughly within the temporal domain of the word. That is, slower movements were ignored, which could occur, for instance, in connection with the re-setting of the head position, which often spanned several words. No distinctions were made between types of directions of movements. However, test annotations revealed that a word may contain either one or two clearly distinguishable beats within a single word and hence, we introduced a distinction between 'simple' and 'double' instances of head or eyebrow movements; we can anticipate that 'double beats' were only

recognized for head, and never for eyebrow movements.

A focal accent was annotated when a rising F0 movement corresponding to the focal H- tone in the Lund model of Swedish prosody (Bruce, 1977; 2007), or to the LH prominence tone in Riad (2006), was recognizable in the F0 contour; note that this F0 movement was expected in the stressed syllable for Accent 1 words, while later in the word, surfacing as a second peak, in Accent 2 words. Praat was used, again, for inspecting F0.

In addition, all 986 words were tagged for lexical pitch accent category. In this very first approach, all words were simply classified as either Accent 1 or Accent 2, according to the lexical or morphological rules of Swedish. That is, it was not actually judged whether a word was, phonetically (non-focally) accented or not. However, as words may be de-accented, which is frequently the case, e.g., in the case of function words, a more detailed analyses of the data will be an important future task.

The analysis in this study is restricted to studying the distribution of head (simple and double movements) and eyebrow movements as a function of word accent type (Accent 1 vs. 2) and prosodic prominence level (focal vs. non-focal). In a first step, a co-occurrence of a focal accent and a movement was counted as such only if an annotation of both events had been made for the same word. However, annotations for focal accents on the one hand, and head- or eyebrow movements on the other, often fell on adjacent words, where it appeared obvious in many of these cases that both events relate to the same word. Therefore, in a second analysis, even such annotations of focal accents and movements on adjacent words were counted as co-occurrences.

Chi-squared tests were used in order to determine whether prominence levels (non-focal, focal) and word accent categories (Accent 1, Accent 2) have a significant effect on the distribution of head and eyebrow movements.

Results

Results of the first and second analysis are displayed in Tables 1 and 2, respectively. Table 1 shows that more than half of the words in the corpus (514) were non-focal Accent 1 words; these include many, probably de-accented, function words. The remaining three accent categories are about evenly distributed in the corpus (146-175 words in each category).

Table 1 further shows that eyebrow and (simple) head movements have been annotated on words of all accent categories, but movements of both types were much more frequent in focally accented words (see also Figure 1): Eyebrow movements were annotated for on average 3.6% of the non-focal words, as opposed to about 14% of the focal words; head movements were annotated for as much as half of the focally-accented words, and again, in far fewer cases for non-focal words.

This effect of the phonological prominence level (focal vs. non-focal) on the distribution of movements proved significant both for eyebrow (χ^2=42.24, p<.01) and head beats (χ^2=209.11, p<.01). In these chi-squared tests, samples for Accent 1 and Accent 2 were collapsed for each prominence category (non-focal, focal); a parallel set of tests was performed separately for Accent 1 and Accent 2 words, resulting in lower, but still significant χ^2 values.

Figure 2 suggests that the word accent type might have some effect on the distribution of,

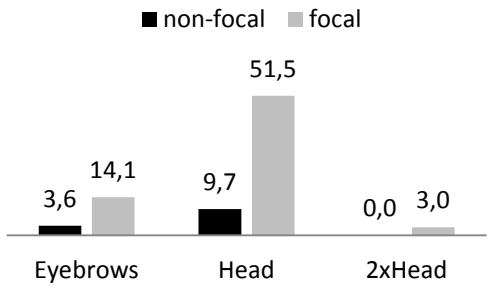

Figure 1. Distribution of eyebrow and head movements in % as a function of phonological prominence level (focal, non-focal), pooled across word accent types.

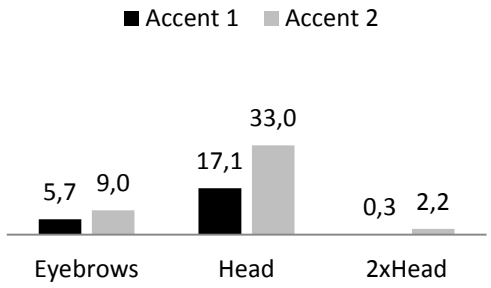

Figure 2. Distribution of eyebrow and head movements in % as a function of word accent category (Accent 1, Accent 2), pooled across prominence levels (focal, non-focal).

Table 1. Distribution of eyebrow and head movements across a corpus of 986 words as a function of phonological prominence level (focal, non-focal) and word accent category (Accent 1, Accent 2). Co-occurrences of focal accent labels and movement labels on exactly the same word.

Movement		Accent category				Total
		non-focal		focal		
		Accent 1	Accent 2	Accent 1	Accent 2	
Eyebrow	**Absent**	**498** (96.9%)	**166** (94.9%)	**129** (85.4%)	**126** (86.3%)	**919**
	Present	**16** (3.1%)	**9** (5.1%)	**22** (14.6%)	**20** (13.7%)	**67**
	Total	514	175	151	146	986
Head	**Absent**	**477** (92.8%)	**145** (82.9%)	**74** (49.0%)	**70** (47.9%)	**766**
	Present	**37** (7.2%)	**30** (17.1%)	**77** (51.0%)	**76** (52.1%)	**220**
	Total	514	175	151	146	986
2x Head	**Absent**	**514** (100%)	**175** (100%)	**149** (98.7%)	**139** (95.2%)	**977**
	Present	**0** (0.0%)	**0** (0.0%)	**2** (1.3%)	**7** (4.8%)	**9**
	Total	514	175	151	146	986

Table 2. As Table 1, but: co-occurrences of focal accent labels and movement labels on the same word or adjacent words.

Movement		Accent category				Total
		non-focal		focal		
		Accent 1	Accent 2	Accent 1	Accent 2	
Eyebrow	**Absent**	**509** (99.0%)	**171** (97.7%)	**121** (80.1%)	**118** (80.8%)	**919**
	Present	**5** (1.0%)	**4** (2.3%)	**30** (19.9%)	**28** (19.2%)	**67**
	Total	514	175	151	146	986
Head	**Absent**	**488** (94.9%)	**154** (88.0%)	**64** (42.4%)	**61** (41.8%)	**767**
	Present	**26** (5.1%)	**21** (12.0%)	**87** (57.6%)	**85** (58.2%)	**219**
	Total	514	175	151	146	986
2x Head	**Absent**	**514** (100%)	**175** (100%)	**149** (98.7%)	**139** (95.2%)	**977**
	Present	**0** (0.0%)	**0** (0.0%)	**2** (1.3%)	**7** (4.8%)	**9**
	Total	514	175	151	146	986

first and foremost, head movements. However, Table 1 shows that this effect is not very consistent, as it is mainly observed for (simple) head movements on non-focal words (and also 2xHead annotations on focal words, see below), and hardly for eyebrow movements.

Accordingly, Chi-squared tests did not reveal any effects of word accent category on eyebrow movements, neither when testing separately for non-focal and focal accents, nor for both prominence levels collapsed (χ^2=3.7, p=.052). However, the last mentioned result is marginally significant. An even clearer, significant effect of word accent was revealed for head movements, both when testing for both prominence levels collapsed (χ^2=31.49, p<.01), and for non-focal accents alone; however, no significant effect of word accent was found for focal accents alone.

This tendency towards fewer (head, and to some degree eyebrow) movements in non-focal Accent 1 than in non-focal Accent 2 words (Table 1, Figure 2), can probably be explained as an artefact of the composition of the corpus: we do find some movements on both non-focal Accent 1 and non-focal Accent 2 words, which might indicate that certain non-focally, but still accented, words indeed attract movements. However, the non-focal Accent 1 sample probably contains many unaccented (function) words, which do not attract as much movement. Therefore, we find relatively fewer movements among non-focal Accent 1 than in non-focal Accent 2 words.

Turning to the '2x Head' annotations, Table 1 and Figures 2-3 show that double head movements were annotated very sparsely, and only for focally accented words. Of the nine annotated items, seven are Accent 2 words. This effect of the word accent category proved significant (χ^2=8.46, p<.01) with non-focal and focal words collapsed. As discussed above for simple head movements, this result could

likewise be explained by the somewhat deviating non-focal Accent 1 sample. However, traces of this effect are still seen when only focal words are included, although not reaching significance (χ^2=3.04, p=.08).

Table 2 displays the results of the second analysis. As mentioned above, in a number of cases, movement annotations did not exactly coincide with words labelled as focally accented, but rather with words directly preceding or following the focal word. In the second analysis, such adjacent movements were also ascribed the focally accented word. Accordingly, as compared to Table 1, Table 2 reveals a certain shift of tokens from the 'Absent' to the 'Present' rows for the focal words, and vice versa for the non-focal words. Note that the '2x Head' annotations are not affected, since these always coincided with focally accented words.

Overall, Chi-squared tests for analysis 2 did not provide any surprisingly different results than those performed for analysis 1.

Discussion

The results have revealed a dependency of the distribution of eyebrow and head movements on the one hand and focal accents on the other, confirming one part of our hypothesis. Furthermore, no (strong) effect of the word accent type on eyebrow and head movements was found; the effect found can probably be explained as an artifact of the composition of the database (see above), which means that the second part of the hypothesis is largely confirmed.

However, there were traces of an effect of the word accent type on the annotations of 'double' head movements. This effect might be explained as follows: Only nine words in the entire corpus were annotated with a double head movement (to be compared with 220 annotations of simple head movements), of which seven were Accent 2 words, all of which were compounds. Compounds are characterized by a complex lexical stress pattern, comprising a main and a secondary stress. In addition, focal Accent 2 words are produced with two pitch peaks. If this effect were corroborated by additional data in future studies, it could imply an association of a head movement, if it is used to add prosodic prominence to a word, with a linguistic/phonetic prominence. Possibly, it requires a lexically stressed syllable to associate with, in a similar manner as is known for accentual tones. A similarly "linguistic" behavior does not seem to be evidenced for eyebrow movements.

A way of interpreting the results by Swerts and Krahmer (2010) is that head movements and eyebrow movements have quite equivalent, cumulative functions as building blocks of prominence, as each of them seems to mirror a minor degree of prominence, while their combination adds up to a higher degree of prominence. In our study, about equally many head movements were annotated (229 of 986 words, Head and 2xHead annotations collapsed) as in the Dutch data in Swerts and Krahmer (228 of 985 words). However, we annotated far fewer eyebrow movements (67 vs. 303). That is, we annotated about two eyebrow movements, on average, per piece of news. This suggests that eyebrow movements were used rather sparsely by the speakers, presumably mostly restricted to words representing the (absolutely) most important information.

This (tentative) conclusion on eyebrow movements, in combination with our (tentative) conclusion drawn above on the "linguistic behavior" of head movements, as well as their relatively frequent occurrence, might suggest that head and eyebrow movements can represent two quite different modalities of prominence cuing, both from a formal and a functional point of view. That is, head movements might have a closer association to low-level prominence and phonological prosodic structure, while eyebrow movements might be more restricted to higher-level prominence and information-structure coding.

The present analysis has provided a preliminary, but significant insight into the usage of head and eyebrow movements in Swedish newsreaders. However, it will need to be further developed, for example by means of an additional classification of the 'non-focal' words into phonetically de-accented and accented words. What we have also neglected so far are co-occurrences of head and eyebrow movements. Future studies will also need to incorporate an analysis of the information-structural conditions underlying the distributions of the movements, as well as a phonetic analysis of the focal and non-focal accents.

A further question for future studies (using the present and additional materials) concerns the timing of movements and focal accents. As an informal note, we have observed a number of instances of very (phonetically and visually) prominent words, which also seem to represent

the informational focus of the message, and which were associated with both a head and an eyebrow movement. In these cases, eyebrow movements (usually raises) often seem to precede the head movement. This could indicate that eyebrow movements can function as a kind of upbeat for a (multimodal) prosodic prominence, which would imply further evidence for the claim that eyebrow and head movements do not simply represent cumulative functions.

Acknowledgements

This work was supported by a grant to the first author from the Marcus and Amalia Wallenberg Foundation.

References

Boersma P, Weenink D (2014). Praat: doing phonetics by computer [Computer program]. http://www.praat.org/

Biau E, Soto-Faraco S (2013). Beat gestures modulate auditory integration in speech perception. *Brain & Language*, 124: 143-152.

Bruce G (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Bruce G (2007). Components of a prosodic typology of Swedish intonation. In: Riad T & Gussenhoven C, eds, *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*. Berlin, 113-146.

Casasanto D (2013). Gesture and language processing. In: Pashler H, ed, *Encyclopedia of the mind.* Los Angeles, London, New Delhi, Singapore, Washington DC: Sage, 372-374.

Flecha-García M L (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech communication*, 52: 542-554.

Lucero C, Zaharchuk H, Casasanto D (2014). Beat gestures facilitate speech production. *Proc. of the 36th Annual Conference of the Cognitive Science Society*, Austin, TX, 898-903.

Riad T (2006). Scandinavian accent typology. In: Viberg Å, ed, *Special issue on Swedish. Sprachtypologie und Universalienforschung (STUF*), 59: 36-55.

Swerts M, Krahmer E (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38: 197-206.

Wang L, Chu M (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, 51: 2847-2855.

Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H (2006). ELAN: a professional framework for multimodality research. *Proc. of LREC 2006, Fifth International Conference on Language Resources and Evaluation.* See also: http://tla.mpi.nl/tools/tla-tools/elan/

Yasinnik Y, Renwick M, Shattuck-Hufnagel S (2004). The timing of speech-accompanied gestures with respect to prosody. *Proc. of From Sound to Sense*, MIT, Cambridge, MA, 97-102.

Transfer of rising tone patterns from Swedish L1 into Spanish L2 – which are the communicative consequences[1]?

Berit Aronsson
Språkstudier, Umeå Universitet

Abstract

Would prosodic transfer from Swedish L1 into Spanish L2 have any effects on communication with native Spanish speakers? The study investigates the contribution to the perceived foreign accent of the L2 prosody displayed by Swedish learners, focusing especially on the role played by rising boundary tones and their pragmatic values. Swedish L1, Spanish L1 and Spanish L2 data, analysed acoustically by Aronsson and Fant (2014), were evaluated in perception experiments with native speakers of Spanish and Swedish and the results show that the intersubjective value associated with a rising boundary tone differs depending on whether the evaluator is a native speaker of Swedish or Spanish, and that the transferred patterns not only contribute to a foreign accent, they are also capable of affecting pragmatic values.

Background

In the request evaluated (a booking of a table over the phone), Aronsson and Fant (2014) identified a phonetic transfer into Spanish L2 of boundary rise patterns produced in Swedish L1. The present study investigates the communicative consequences of such transfer. The framework of analysis applied to interpret the results of the perception experiments, also used by Aronsson and Fant (2014), is the so-called Intersubjectivity Management Model, initially proposed by Fant (2006) and Fant and Harvey (2008) for conversational analysis. Additionally, the framework proposed by Brown and Levinson (1987) for the understanding of politeness strategies has been used to interpret some aspects of the results. The Intersubjectivity Management Model enables the separation of two different intersubjective values associated with the rising boundary tones in interactional speech, namely transactional (+/- request for information), and interpersonal (+/- friendliness/politeness)[2] values. The results presented by Aronsson and Fant (2014) showed that the L2-speakers did not master the tonal differences between continuative tones and question patterns produced in Spanish L1, as used in the opening and closing unit of the request (Figure 1-2). The rise types also differed phonetically in the lengthening of the final vowels produced at the end of the rise (irrespective of whether these vowels were stressed or not (Figure 3). The rise patterns found were also produced in contexts where L1 Spanish speakers seemed to prefer falls.

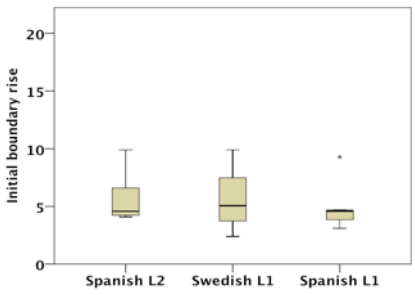

Figure 1 (cited from Aronsson and Fant 2014). Boxplot illustrating group variation in opening units.

[1] A more elaborated version of this paper is to be published in Spanish in *Onomázein,* Aronsson, Berit (forthcoming)

[2] From the point of view of politeness theory (Brown and Levinson 1986), friendliness could be seen as related to positive politeness, an act aimed to establish a positive relationship between the interlocutors

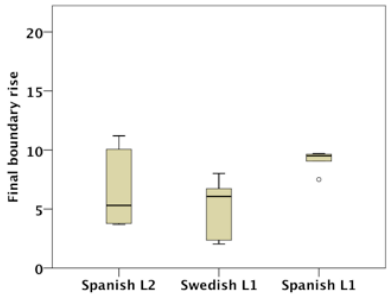

Figure 2 (cited from Aronsson and Fant 2014). Boxplot illustrating group variation in closing units.

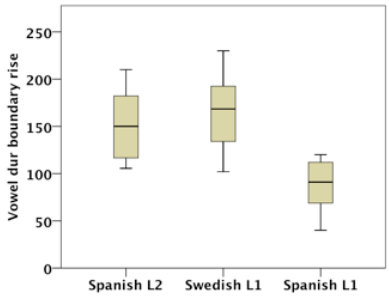

Figure 3 (cited from Aronsson and Fant 2014). Boxplot illustrating group variation for unit-final vowel duration in rises.

The underlying reasons for the transfer observed in Aronsson and Fant (2014) can be discussed in terms of different intersubjectivity-related values associated with the tonal rises:

Differences in discursive strategies between Spanish and Swedish, where open-ended 'yes'-'no'-questions are by default characterised by a high final rise in (Peninsular and Chilean) Spanish (Navarro 1944, Quilis 1985, 1993, Sosa 1999, Román et al. 2008), but only optionally produced in Swedish in these contexts (Bredvad-Jensen 1984, Gårding 1998, Elert 2000, Ambrazaitis 2009), were proposed as explanations to this transfer. A friendliness– rather than a purely information-seeking value, associated with final rises in both wh-questions and open-ended word-order questions (also labelled 'yes'-'no' questions), has been reported by Kohler (2004) for German, and by House (2005) for Swedish. This value, as used in Swedish, by no means seems to be limited to questions, since also friendly declaratives tend to end in rises (as suggested by Hadding Koch and Studdert Kennedy 1969: 176, and discussed in terms of intersubjectivity management by Aronsson and Fant 2014).

In Spanish on the other hand, a terminal fall has instead been associated with politeness in some interrogative types (Font-Rotchés and Mateo, 2013: 269). In Catalán, a language typologically similar to Spanish and spoken in the region of Cataluña, Spain, an increased F0 in a terminal boundary rise in yes-no-questions correlated with a lower degree of perceived politeness[3] (Nadeu and Prieto, 2011: 850).

Based on these findings, the paper suggests that the fundamental value of a boundary final rise in the task evaluated is that of friendliness/positive politeness in Swedish, while a final rise produced in a corresponding L1 Spanish context is mainly information-seeking.

Aims

The main aim of the present study is to investigate possible differences in the degrees of the perceived transactional and interpersonal intersubjective values associated with a boundary fall/ rise produced in a spontaneous request in Spanish L1, Spanish L2 and Swedish L1. In order to study these values, two perception experiments with 27 native speakers of Spanish from Chile and Spain and 34 Swedish native controls were carried through. In a pre-test the study investigates the contribution of Spanish L2 boundary tone realisation to foreign accent.

Procedure

As in Aronsson and Fant (2014) the authenticity of the dialogue has been considered an important prerequisite, and the same spontaneous task studied acoustically in Aronsson and Fant (2014) is evaluated in the present study by native speakers.

A pre-test, which assessed the possible contribution of boundary tones produced in Spanish L2 to the foreign accent perceived by the native speakers, was initially carried out to justify (or not) a further investigation of the values associated with rising boundary tones. Experiment 1 assessed the transactional (=information seeking-) and interpersonal (=friendliness-) values associated with a rising and falling boundary tone respectively, extracted from the data set and presented as tone alone without the segmental information. Experiment 2 examined whether the perceived interpersonal

[3] 'Politeness' is defined in Nadey and Prieto (2011) basically according to Ohala's (1986) framework, where rising F0 is believed to correlate with values such as 'polite' and 'friendly'.

value (=perceived friendliness) associated with these tones differed between the groups when presented in a non-manipulated, contextualized form, in L1 and L2 Spanish and L1 Swedish. The falls and rises evaluated in L1 and L2 Spanish and L1 Swedish displayed similar tonal ranges but were characterized by a difference in duration of the final vowel produced at the end of the boundary tone, which was shorter in L1 Spanish.

Results

Pre-test

The results demonstrated that, regardless of the variety of Spanish used by the evaluator (Chilean or peninsular Spanish), the rising boundary tone (RBT) was the suprasegmental feature that most characterized the Spanish foreign accent of Swedish speakers. Stress placement (SP) and vowel lengthening (VL) came in second place. These categories, evaluated in the pre-test by 27 Spanish L1 speakers, were based on a preliminary initial experiment, carried through prior to the pre-test: 10 native Spanish L1 speakers were recorded while imitating and describing how the foreign accent produced in L2 Spanish sounded to them, these were not the same as in the pre-test. The subjects were asked to describe the foreign accent perceived only by reacting to what they heard, i.e. according to their previous (emotional) priming related to this kind of request, without being asked to specify whether segmental or suprasegmental features were involved and without knowing our objective. Based on these recordings the alternatives used in the pre-test were formulated and categorised into segmental and suprasegmental features respectively.

This method implies that the alternatives formulated are not phonetically precise. It would for example be possible that the final lengthening of the vowel at the end of a rising boundary tone was a contributing factor to the rise being interpreted as "foreign accented", even though the subject him/herself failed to recognise this (untrained listeners probably perceive whole chunks rather than isolated features). In fact, in rises with similar rise range in Spanish L1 and Spanish L2, this is the only acoustically significant suprasegmental difference identified between the Spanish L1 and L2 data studied, a finding that will be further addressed below (in Experiment 2).

Experiment 1

The results of Experiment 1 showed significant differences between the groups in the interpretation of the information seeking value perceived in the rising boundary tone: the subjects from Chile and Spain always interpreted this tone as a question while this was not always the case in the Swedish group: 30% of the Swedish subjects perceived a descending tone as an interrogative pattern (Figure 4).

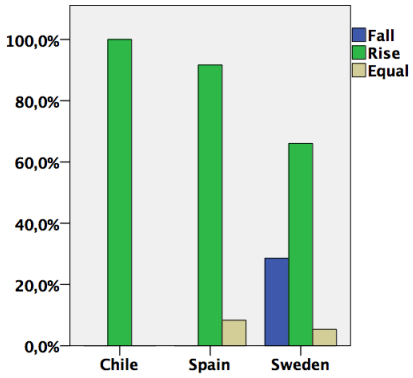

Figure 4. Distribution of the values "fall= question", "rise=question" perceived by Chilean, Spanish and Swedish speakers.

The Swedish and Chilean group perceived a rising tone as friendlier than a descending one to a larger degree than the peninsular Spanish group did (Figure 5). The explanation I propose for this is that the speakers of peninsular Spanish perceived this pattern to be too invasive, a threat to their negative face, which is the reason why the descending pattern was preferred. It seems there is greater tolerance towards positive politeness in the Chilean and Swedish groups than in the peninsular Spanish group, whereas the latter expressed greater preference for the interlocutor not to invade their 'private territory'. The Chilean and Swedish groups obtained very similar results, since they perceived the rising pattern as friendlier. However, it seems that the underlying reasons have their origin in a divergent definition of what is perceived as 'friendly': The Chilean group perceives the rising pattern to be friendlier than the descending one since it interprets the rising tone as a question, which is not so clear for the Swedish group. For the Swedish subjects the value of friendliness seems to be linked to

the rising tone itself, regardless of whether or not it is interpreted as a question.

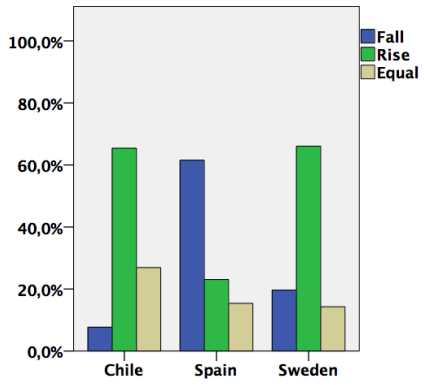

Figure 5. Distribution of the values "fall= friendly", "rise= friendly" perceived by Chilean, Spanish and Swedish speakers.

Experiment 2

Utterances of the same pragmatic force were used (a greeting and the final unit of the request act), and units of similar rise range were tested. What differed, however, was the duration of the final vowel of the rise/fall (longer in Spanish L2/Swedish L1). The trend observed in Experiment 2 was that, in the contextualized speech of the current task, a majority of the Spanish native speakers, both from Chile and from Spain, perceived a rising tone produced in L1 Spanish as friendlier than a falling tone, while the rising tone produced in L2 Spanish or L1 Swedish on the other hand tends to be evaluated as less friendly than the fall (Figure 6-7).

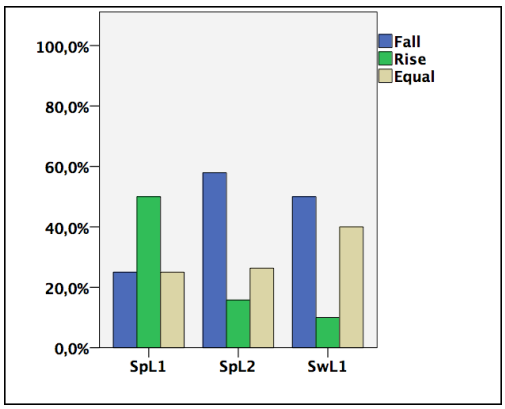

Figure 6. Speakers of Chilean Spanish: Distribution of the values "fall= friendly", "rise= friendly" in opening units produced in Spanish L1, L2 and Swedish L1.

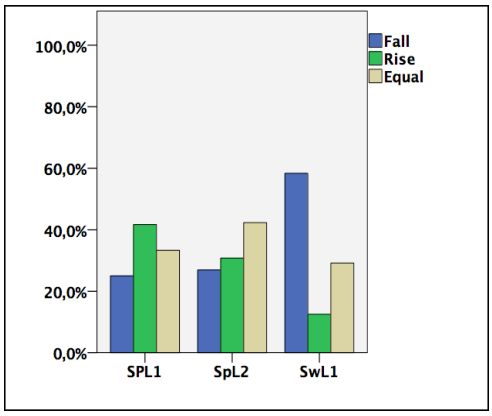

Figure 7. Speakers of Peninsular Spanish: Distribution of the values "fall= friendly", "rise= friendly" in opening units produced in Spanish L1, L2 and Swedish L1.

This suggests that a Swedish speaker, transferring tonal patterns from the mother tongue to Spanish, risks being perceived as less friendly than a native speaker, even if in both cases the pattern were rising. Since the acoustic difference between the rising tones of L1 and L2 Spanish and L1 Swedish lies mainly in a difference in the final rising vowel (longer in L2 Spanish/L1 Swedish), we attribute this result to the greater vowel lengthening identified in L2 Spanish/L1 Swedish than in L1 Spanish.

Conclusions

The result of the pre-test showed that the characteristics of the rising patterns in the boundary tones are not only *performed* differently from the native realizations (Aronsson and Fant 2014), they are also *perceived* as different. The foreign accent identified seems to be associated mainly (although by no means entirely) with the realisation of the rising boundary tones. Different interpretations of the intersubjective values associated with these rises are discussed as underlying explanations to the foreign accent perceived by the Spanish L1 speakers, where a rising boundary tone is interpreted as information seeking in Spanish L1 but tends to be associated with friendliness in Swedish L1. The paper suggests that the significantly longer final vowels of the rising tones produced in Swedish L1 and Spanish L2 may have increased the Spanish L1 speakers' feelings of having their private territory invaded; they might therefore have perceived them as a greater threat to face than the descending tone. Several additional

explanations are also possible, as for example less tolerance for invasive patterns when the acoustic pattern doesn't correspond to the expected one, i.e. when it is produced by an L2 speaker. It should finally be remembered that the samples studied are small, which implies that the results should be taken with certain precaution. The values associated with the boundary tones and the communicative consequences of this type of tonal transfer should be further tested in future studies.

References

Ambrazaitis, Gilbert (2009). Nuclear intonation in Swedish. Evidence from Experimental- Phonetic Studies and a Comparison with German. Doctoral dissertation. Lund University, Department of Linguistics.

Aronsson, Berit & Lars Fant (2014). Boundary Tones in Non-Native speech: the Transfer of Pragmatic Strategies from L1 Swedish into L2 Spanish. Intercultural Pragmatics, 12(2), 159-198.

Bredvad-Jensen, Anne-Christine (1984). Tonal geography. Geographical variation in declarative and interrogative intonation along the west cost of Sweden. In Elert, Claes-Christian, Johansson, Irene & Strangert Eva (eds). Nordic Prosody III, 31-41. University of Umeå.

Brown, Penelope & Stephen C. Levinson (1987) [1978]: Politeness. Some Universals in Language Use, New York: Cambridge University Press.

Elert, Claes-Christian (2000). *Allmän och svensk fonetik* (8th ed.). Stockholm: Norstedts.

Fant, Lars (2005). La entonación: informatividad, emotividad, dialogicidad. *Filología y Lingüística. Estudios ofrecidos a Antonio Quilis*. *Volumen I,* 191–218. Madrid: Consejo Superior de Investigaciones Científicas.

Fant, Lars & Ana María Harvey (2008). Intersubjetividad y consenso en el diálogo: análisis de un episodio de trabajo en grupo estudiantil, *Oralia, Análisis del discurso oral* 11, 307–332.

Font-Rotchés, Dolors & Miguel Mateo Ruiz (2013). Entonación de las interrogativas absolutas del español peninsular del sur en habla espontánea. *Onomazéin* (28), 256-275.

Gårding, Eva (1998). Intonation in Swedish. In Hirst, Daniel, Di Cristo, A. (Eds.), *Intonation Systems.* Cambridge University Press, Cambridge, 112–130.

Hadding-Koch, Kerstin & Michael Studdert-Kennedy (1964). An experimental study of some intonation contours. *Phonetica* 11. 175-185.

House David (2005). Phrase-final rises as a prosodic feature in wh-questions in Swedish human-machine dialogue. *Speech Communication* 46. 268-283.

Kohler, Klaus (2004). Pragmatic and Attitudinal Meanings of Pitch Patterns in German Syntactically Marked Questions. In Fant, Gunnar, Fujisaki, Hiroya, Cao, Jianfen, Yi Xu, (Eds.*), From Traditional Phonology to Modern Speech Processing. Foreign Language Teaching and Research Press*, Beijing, 205–214.

Nadeu, Marianna & Pilar Prieto (2011). Pitch range, gestural information, and perceived politeness in Catalan, *Journal of Pragmatics* 43 (3), 841-854.

Navarro Tomás, Tomás (1944). *Manual de entonación española*. New York: Hispanic Institute in the United States.

Ohala, John (1984). An ethological perspective on Common Cross-Language Utilization of F0 of voice, *Phonetica* 41, 1-16.

Quilis, Antonio (1985). "Entonación dialectal hispánica", *Lingüística Española Actual* VII, 145-190.

Quilis, Antonio (1999) [1993]: Tratado de fonología y fonética españolas, Madrid: Gredos.

Román, Domingo Montes de Oca, Valeria Cofré Vergara & Claudia Rosas Aguilar (2008). Rasgos prosódicos de oraciones sin expansión, del español de Santiago de Chile en habla femenina, *Language design. Special issue* 2, 137-146.

Sosa, Juan Manuel (1999). *La entonación del español*. Madrid: Cátedra.

The realisation of sj- and tj-sounds in Estonian Swedish: some preliminary results

Eva Liina Asu[1], Otto Ewald[2] and Susanne Schötz[3]
[1]Institute of Estonian and General Linguistics, University of Tartu
[2]Centre for Languages and Literature, Lund University
[3]Department of Logopedics, Phoniatrics and Audiology, Lund University

Abstract

The aim of the present study is to cast some light on the realisation of sj- and tj-sounds (/ɧ/ and /ɕ/) in Estonian Swedish. These sounds are said not to occur in this variety of Swedish, or to have a different realisation as compared to Standard Swedish (Lagman, 1979). Our analysis, based on repetitions of elicited words produced by five elderly speakers of Estonian Swedish, shows that sj- and tj-sounds are mainly pronounced the way they are spelt (e.g. tjock [tjɔk:] (thick, fat) or stjärna [stja:ɳ] (star)). There appears, however, some variation in the realisation of the sj-sound, which depending on the word and position in the word can also be pronounced as a retroflex fricative [ʂ] or a palatal fricative [ç]or [ɕ].

Introduction

This study presents a small-scale investigation of sj- and tj-sounds in Estonian Swedish – a highly endangered variety of Swedish that until WWII was spoken on the islands and North-Western coastal areas of Estonia, but is currently surviving only in the speech of a community of elderly emigrants to Sweden, and a handful of speakers in Estonia.

The acoustic characteristics of the Estonian Swedish sound system are still very little studied (except for a study of close vowels by Asu et al. (2009), and two studies of the lateral fricative: Schötz et al. (2014) and Asu et al. (2015)). The descriptions found in the older accounts of dialectal research are often rather general or even hard to interpret. According to Lagman (1979) tj- and Standard Swedish sj-sound do not occur in Estonian Swedish; thus, the word *tjäna* (to earn, to serve) would be pronounced with a /t/ followed by what he describes as an apico-alveolar fricative *sj* (*spetsigt sj-ljud*), and the word *sjö* (lake) would be pronounced with the apico-alveolar fricative (Lagman, 1979:11). This description is intriguing, as on the one hand it claims that the tj-sound is absent from the phonological inventory, whereas on the other hand it suggests that the Estonian Swedish realisation of this sound is an affricate (which is a common realisation e.g. in Finland Swedish (Garlén, 1988)). Lagman's description also suggests that the Estonian Swedish sj-sound is realised differently from Standard Swedish.

For the Swedish sj-sound, which has over 30 different spellings (Garlén, 1988), two main variants are distinguished: [ɧ] (or [ɧ^{t}]) and [ʂ] (or [ʃ]) which are respectively referred to as 'dark' (the dorsal velar variant, commonly used in the Southern Standard Swedish), and 'light' (the predorsal apical variant, used mainly in the northern varieties) (Elert, 2000; Malmberg, 1971). Next to these variants also [ɕ] occurs mainly in Finland Swedish (Leinonen, 2004: 60). There are, however, many more allophones along the continuum between the two main variants depending on the regional, cultural and social aspects, as well as the age and sex of the speakers (Lindblad, 1980). The so called 'light' variant [ʂ] is sometimes considered to have higher prestige, occurring more often in the speech of women and higher social classes (Bruce, 2010: 166) and also read speech (Lindblad, 1980: 10); [ʂ] is more common among older generations and [ɧ] among younger ones.

Likewise, the orthography, phonetic realisation and transcription of the tj-sound varies, although not as widely as for the sj-sound. The most common spellings are *tj, k* and *kj*. The prevailing realisation, typical of the Central Swedish standard variety, is an alveolo-palatal fricative [ɕ], which in the older literature has also been transcribed with the IPA symbol for the dorso-(pre)palatal sibilant [ç] (e.g. Elert,

2000). In some areas of Sweden (South-Eastern and Northern Sweden) the tj-sound can be realised as a prepalatal affricate [cɕ], which occurs more frequently in the speech of elderly speakers (Elert, 2000). An affricate [t͡ɕ] or [t͡ʃ] is also a common realisation in Finland Swedish (Garlén, 1988: 71).

The Swedish sj- and tj-sounds, and fricatives on the whole, are notoriously difficult to perceive and describe acoustically (e.g. Lindblad, 1980). In addition to huge dialectal differences this is due to the large interspeaker variation in the realisation of these sounds, which has been explained by the varying shapes and sizes of the speakers' oral cavity and front tongue (e.g. an EPG analysis by Lindblad and Lundqvist (1995)).

The aim of the current study is to analyse the phonetic realisation and variation in the pronunciation of sj- and tj-sounds in Estonian Swedish using a set of elicited isolated words. Following Lagman (1979) we would expect words with these sounds to be pronounced according to their orthography; the tj-sound could also be realised as an affricate.

Materials and method

Speakers and speech data

The data was recorded in 2009 in Stockholm from four elderly speakers of Estonian Swedish (2 women and 2 men), and in 2012 in Nuckö (Noarootsi), Estonia, from one elderly female. All speakers represent the Nuckö-Rickul variety – the largest dialectal area of Estonian Swedish. The speakers recorded in Stockholm had arrived in Sweden in the mid 1940s as youngsters and were between 80 and 86 years old at the time of the recording. They were recorded in a quiet setting in Stockholm using a Sony portable DAT recorder TCD-D8 and Sony tie-pin type condenser microphones ECM-T140. The speaker recorded in Estonia was 77 years old at the time of the recording. The recording was conducted at her home using a Roland R-09HR WAVE/MP3 recorder with a Sony tie-pin type condenser microphone ECM-T140.

As materials, a word list adapted from the word list of the SweDia 2000 database was used (slightly different word lists were used in the two recording sessions). The words were not read but elicited, and the subjects were asked to repeat each word at least three times. The data recorded in Stockholm consists of seven words where /ɧ/ or /ɕ/ appear word-initially in other varieties of Swedish (four words with /ɧ/ and three with /ɕ/: *chaufför* (chauffeur), *sjö* (lake), *stjärna* (star), *skjorta* (shirt), *körhjul* (bicycle), *kälke* (sleigh), *tjock* (thick, fat)), and the data from Nuckö comprises six such words including five with /ɧ/ and one with /ɕ/ (*ske/r* (happen/s), *sjunga* (sing), *sjö* (lake), *stjärna* (star), *skjorta* (shirt), *tjuv* (thief)). Additionally, the dataset recorded in Stockholm included three words where variants of /ɧ/ would appear word-finally or medially in Estonian Swedish: *lös* (loose) pronounced with a diphthong [au], *Pauls (gård)* (Paul's (farm)) pronounced as a disyllabic word, and *nors* (European smelt, a species of fish), an example of supradental assimilation.

For a comparison with Central Swedish the word list data from the SweDia 2000 database (Bruce et al., 1999) was used. The data comprised three repetitions of two words containing /ɧ/ and /ɕ/: *själen* (the soul) and *käke* (jaw), recorded in a quiet home setting with a Sony portable DAT recorder TCD-D8 and Sony tie-pin type condenser microphones ECM-T140. Three elderly women and three elderly men from the location Kårsta near Stockholm were selected from the SweDia 2000 database. The speakers were between 64 and 74 years old (mean age 67).

Analysis

The data was manually labelled, segmented, normalised for intensity, and analysed using Praat (Boersma & Weenink, 2015). All realisations of /ɧ/ and /ɕ/ were first classified based on auditory analysis and visual examination of spectrograms.

Three measures were taken of all the sibilant sounds: duration, mean relative intensity and the centre of gravity (COG) of the DFT spectrum (the so-called 'spectral centroid'). Duration and mean relative intensity were obtained using a Praat script. In order to measure the centre of gravity, the steady state part of the sibilants was first manually segmented, which usually meant that about 10-15 ms were excluded from the beginning as well as from the end of the sound to minimise the influence of coarticulation. Then a Praat script was run. Mean values were calculated for all the measures for the different sibilant realisations in different words.

Results and discussion

Phonetic realisation of sj- and tj-sounds

Table 1 presents the different realisations of sj- and tj-sounds in the test-words. It appears (as expected) that the tj-sound is always realised the way it is spelt. Thus, in the words *tjuv* and *tjock* it is pronounced with [tj], and in the words *körhjul* and *kälke* it is realised without velar softening as [k] for all speakers. Contrary to our expectations there were no affricate realisations in the data.

Table 1. The phonetic realisation of sj- and tj-sounds in the Estonian Swedish data. The total number of tokens for each word is shown in the brackets.

Test word	Phonetic realisation
chaufför (12)	ʂ
sjö (16)	sj, ç, ɕ
stjärna (13)	stj
skjorta (14)	skj
ske/r (3)	sk
sjunga (3)	sj
lös (21)	ʂ
nors (12)	ʂ
Pauls (15)	ʂ
tjock (12)	tj
tjuv (3)	tj
körhjul (9)	k
kälke (15)	k

For the sj-sound there was more variation in the productions. In most words with an initial sj-sound, the pronunciation matched the orthography: *stjärna* was pronounced with [stj] and *skjorta* with [skj] for all speakers. Also, the word *ske/r* was realised with [sk] and *sjunga* with [sj].

The sibilant in the word *chaufför* was in all cases produced as a retroflex fricative [ʂ]. The realisations of the word *sjö* varied somewhat: in addition to the most common [sj] (10 tokens), also a palatal fricative [ç] and an alveolo-palatal fricative [ɕ] appeared in the pronunciation of two speakers. The palatal fricative realisation might seem surprising at first glance, but becomes more logical if viewed as a variant of [sj] that has taken the place of articulation from the second element [j] and the voicing from the first element [s]. More speakers and comparable test words are needed to see if this indeed is the case rather than merely a speaker-specific variation, or if the alveolo-palatal fricative variant similar to the realisation of sj-sound in Finland Swedish (cf. Leinonen 2004) is more common also in Estonian Swedish.

In addition to word-initial sibilants the data of the four speakers included three words where the fricative appeared word-finally. In all these instances it was produced with an [ʂ]: *lös* [l͡auʃ], *Pauls* [poːʃa] and *nors* [nɔʂ]. The latter is an example of supradentalisation (retroflexion), i.e. the assimilation of /r/ and /s/ into [ʂ] found in most Swedish dialects (except in South and Finland Swedish). It is interesting that Estonian Swedish differs in this respect from Finland Swedish.

Acoustic measures

Table 2 presents the average measures of duration, mean intensity and COG for the different variants of fricative sounds found in the data. The figures should be interpreted with caution, as the total number of tokens is small. Nevertheless, some general observations can be made on the basis of these results. The best measure for differentiating between the various fricative sounds seems to be the centre of gravity. Likewise, in a comparison of [s] and [ʃ], Nittrouer et al. (1989) found that COG reliably distinguished between the spectral shapes of the two fricatives, yielding higher values for [s] than for [ʃ].

Table 2. Average measures of duration (ms), intensity (dB) and centre of gravity (COG) (Hz) for the different realisations of sj- and tj-sounds in the Estonian Swedish and Central Swedish data.

Sound	Duration	Intensity	COG
Estonian Swedish			
s	128	52	5007
ɕ	120	65	4732
ç	116	46	1692
ʂ (chaufför)	100	53	2979
ʂ (nors)	300	62	4804
ʂ (lös)	198	60	4832
ʂ (Pauls)	129	61	3837
Central Swedish			
ɧ	113	65	933
ɕ	119	62	3991

In our data COG was highest for [s] measured at the onset of consonant clusters *sj, stj, skj, sk* (average 5007 Hz), and lowest for [ç] (1692 Hz). The COG for [ʂ] was on average 2979 Hz in the word *chaufför*, but had a higher value word-finally: 4832 Hz in *lös* and 4804 Hz in *nors*, and 3837 Hz in *Pauls*. As a comparison, the sj-sound measured in the data of Central Swedish speakers' productions of the word *själen* had a very low COG (933 Hz), which is characteristic of velar sounds, while [ɕ] in the word *käke* had a much higher COG (3991 Hz).

Duration has been shown not to be differentiating for word-initial sibilants (Lindblad, 1980), and in our data all word-initial sibilants had an average duration of 100-128 ms. The word-final sibilants were, however, much longer with [ʂ] in *nors* being the longest (300 ms on average).

Conclusions

The present study analysed the realisation of sj- and tj-sounds in Estonian Swedish elicited words. The sj- and tj-sounds were mainly pronounced according to their orthography. There was variation in the pronunciation of *sj* in the word *sjö*, where in additioan to the prevailing realisation [sj] also the palatal fricative realisations [ç] and [ɕ] occurred.

As the present data set is rather limited the next obvious step is to collect more Estonian Swedish data that would enable us to carry out larger-scale acoustic measurements of these fricatives. We plan to compare sj- and tj-realisations with data from Finland Swedish.

Acknowledgements

The work on this paper was supported by the project Estonian Swedish Language Structure (ESST) (Swedish Research Council, grant no. 2012-907) and Estonian Research Council grant IUT2-37. We would also like to thank Francis Nolan for insightful comments on the topic, as well as our Estonian Swedish informants in Sweden and in Estonia, and Svenska Odlingens Vänner in Stockholm.

References

Asu E L, Schötz S, Kügler F (2009). The acoustics of Estonian Swedish long close vowels as compared to Central Swedish and Finland Swedish. *Proceedings of Fonetik 2009*, Dept. of Linguistics, Stockholm University, 54-59.

Asu E L, Nolan F, Schötz S (2015). A comparative study of Estonian Swedish voiceless laterals: are voiceless approximants fricatives? *Proceedings of the ICPhS 2015*, Glasgow.

Boersma P, Weenink D (2015). Praat: doing phonetics by computer [Computer program] Ver. 5.4.01, retrieved from www.praat.org/.

Bruce G, Elert C-C, Engstrand O, Eriksson A (1999). Phonetics and phonology of the Swedish dialects – a project presentation and a data- base demonstrator. *Proceedings of ICPhS 99* (San Francisco), 321-324.

Elert C-C (2000). Allmän och svensk fonetik. Stockholm: Norstedts.

Garlén C (1988). *Svenskans fonologi*. Lund: Studentlitteratur.

Ladefoged P, Maddieson I (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lagman E (1979). *En bok om Estlands svenskar. Estlandssvenskarnas språkförhållanden*, Vol. 3A. Stockholm: Kulturföreningen Svenska Odlingens Vänner.

Leinonen K (2004). *Finlandssvenskt sje-, tje- och s-ljud i kontrastiv belysning.* Jyväskylä: University of Jyväskylä.

Lindblad P (1980). *Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv*. Travaux de l institut de linguistique de Lund XVI. Lund: Gleerup.

Lindblad P, Lundqvist S (1995). The groove production of Swedish sibilants - an EPG analysis. *Proceedings of the XIIIth International Congress of Phonetic Sciences* 95, 2: 458-461.

Malmberg B (1971). *Svensk fonetik.* Lund: Gleerups.

Nittrouer S, Studdert-Kennedy M, McGowan R S (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults, *Journal of Speech and Hearing Research*, 32: 120-132.

Schötz S, Nolan F, Asu E L (2014). An acoustic study of the Estonian Swedish lateral [ɬ]. In *Proceedings of Fonetik 2014*, Department of Linguistics, Stockholm University, 23-28.

Grimaldi's "Discovery of the Cat Language": A theory in need of revival (or perhaps not?)

Robert Eklund
Department of Culture and Communication, Division of Language and Culture, Linköping University

Abstract
Recent years have seen a growing number of studies on both felid vocalizations in general and human–felid communication in particular. Frequently considered as the starting point for this line of research is Mildred Moelk's seminal paper from 1944, in which she provides a taxonomy of basic felid vocalizations, complete with phonetic transcriptions. Less known is the fact that Cat Language was decoded in far more detail half a century earlier, by one "Prof. Grimaldi", who sadly never published his findings. However, an English translation of Grimaldi's findings was published by Marvin Clark in 1895, so the astonishing observations made by Grimaldi are not lost to the world. In the present paper a summary of Grimaldi's results will be provided, in the hope that this research will serve as a source of inspiration to present and future researchers of Cat Language.

Introduction

Cats have lived with humans for thousands of years, and are (most likely) alongside the dog our oldest companions of another species. Thus, it comes as no surprise that research has been devoted to the decoding of the "cat code" (e.g. Schötz, 2013), and how this might help us humans communicate with our companions.

What is less known is that the 'cat language' (CL) was given an ambitious, exhaustive, exhilarating, mind-blowing – and also somewhat controversial – description more than 100 years ago by one "Prof. Grimaldi", who sadly never published his findings in a scientific journal.

Fortunately, an English translation of Grimaldi's findings was published by Marvin Clark in 1895, so the astonishing observations made by Grimaldi are not lost to the world. However, despite the fact that Grimaldi's observations have existed in print for more than a century, this seems to have been largely ignored in the literature, and with the exception of Grown (2014: 82) mentions of Grimaldi seem to be nonexistent.

In this paper a short summary of Grimaldi's results will be provided, in the hope that this research will serve as a source of inspiration to future researchers of Cat Language.

Background: Clark meets Grimaldi

In the year 1895 a certain Marvin R. Clark published a book entitled "PUSSY and Her Language". Clark was (he tells us) at the time "editor of a New York morning newspaper" (p. 43; all subsequent page numbers will refer to Clark, 1895). Clark's book is singing the praise of the cat in general, and includes mentions about what famous and historical people loved cats, how the cat was revered in ancient Egypt and so on and so forth.

However, the main part of the book focuses on the linguistic capabilities of the cat. The language part, however, was not written by Clark himself, but by a "French gentleman of about fifty years of age" (p. 43) Clark had previously received the gentleman's card, which read (p. 43):

> "Alphonse Leon Grimaldi, F. R. S., F.G. S., M.O.
> D. H. de C., M. F. A. S., M. F. A., et al.
> "Rue du Honore, 13, Paris.
> "Metropolitan Hotel, N.Y."

To make a fairly long story short, Clark and Grimaldi met (they communicated in French, since Clark's mastery of the French was "much more comprehensible" (p. 44) than Grimaldi's English, Clark thought) and during the course of their conversation, Grimaldi revealed to Clark that he (Grimaldi) had "made a life study of the animal kingdom" (p. 44) and presented Clark with a paper that he had written on "Cat language" (p. 44). Grimaldi explained that he had not dared publish his findings, since "he never could have lived through the sarcasm and taunts of those men of science, who would have over-whelmed him with abuse" (p. 45).

The men bid farewell, and Clark forgot about the paper given to him. Years later, however his "meory [sic!] recurred to it", (p. 45) and he was happy to find it intact. When Clark (finally) read the paper, he "rejoiced /… / because it verifies my own [i.e. Clark's; RE] theories, and proves beyond a doubt that the Cat has a language which may be spoken by anybody who will make a study of it" (p. 45). Thus, Clark "made a literal translation" (p. 46) of Grimaldi's paper, with the title "THE CAT", which appears on pages 46 through 122, and whose central findings will be summarized in the following sections.

Anatomy and physiology

Not inclined to leave any aspect or consideration unturned, Grimaldi provides an overview of the cat's anatomy and physiology, with comparisons with other animals. Some of his Grimaldi's main points are summarized below.

General aspects

Grimaldi devotes a passage to the cat's organism in general, which more or less in toto sounds like a summary: "Anatomists are unanimous in their opinions and their experiments show conclusively that the Cat has a much finer and more delicate organism than the dog" (p. 66). Moreover, "it is almost universally conceded that Cats are fully as intelligent as dogs, and by many the feline is regarded as the superior animal in every respect" (p. 66).

Neurological aspects

Not only is the cat superior to the dog in general, when studying the feline brain, more surprises are to be found: "The brain of the Cat so closely resembles that of man as to force the unwilling admission from anatomists and physiologists that in form and substance they bear so close and striking a similarity that they are, to all intents and purposes, the same in substance and conformation, and differ only in weight and size" (p. 65). So the inevitable conclusion is, not surprisingly, that: "the intelligence of the Cat is equal to that of man" (p. 65). Modern researchers and research would most likely only agree with Grimaldi up to a point on that one.

Vocal organs

Grimaldi first establishes that some "mammalia, such as the giraffe, the porcupine, and the armadillo, have no vocal chords [sic!], and are therefore, mute" (p. 64). The cat (or "Cat", which Grimaldi obviously prefers), however, not only have vocal cords (or "chords", as Grimaldi prefers to spell it), but the vocal organism is treated in some detail: "Cats have a sac between the thyroid cartilage and the oshyiodeum, which have much to do with the modifying and increasing of the tones of the voice. The laryngeal sacs are small /… / The epiglottis is comparatively small, and there are proportionately small cavities in the thyroid cartilage and the oshyiodeum, which communicate with the ventricles of the larynx and the laryngea-pharyngeal sacs, which give the peculiar softness of musical tone to the feline /… / one of the most delightful characteristics of the Cat" (pp. 64–65).

Of Signs and Sounds

Grimaldi emphasizes that although CL as an oral and vocal language is much richer than humans usually think, cats preponderatingly express themselves by 'Signs', to be described in some detail in a coming section of this paper.

A linguistic description of CL

The vocal/oral language of the Cat will be summarized in the following sections.

Vowels and consonants

Grimaldi provides an insightful and detailed account of the phoneme inventory in CL, and observes that: "Consonants are daintily used, while a wide berth is given to explosives and the liquid letters 'l' and 'r' enter into the great majority of sounds. The sounds of the labials are not frequently heard, but the vowels, a, e, i, o and u, go far toward making up the entire complement of words in the language" (p. 103).

Of interest is that Grimaldi almost touches upon aspect theory when he mentions that certain phonemes go hand in hand with certain moods. For example, he mentions that the "sounds of the labials, b, f, m, p, v, w, and y, are more frequently heard in words of anger than otherwise" (p 106). As an example he mentions the cat's "significant war-cry /…/ mie-ouw, vow, wow teiow yow tiow, wow yow, ts-s-s-s-syow" (p. 106), an expression so bold that Grimaldi refrains from putting it in English. However, given that the word "yow" means "extermination from the face of the earth" (p. 106) one can easily guess what the above utterance is meant to signal.

In fairness, Grimaldi might have been inspired by another Frenchman, Champfleury, who had already pointed out that Cat Language possesses the same vowels as Dog Language, but that CL also includes "six consonnes : l'*m*, l'*n*, le *g*, l'*h*, le *v* & l'*f*" (Champfleury, 1869, p. 191). This, however, must remain mere speculation.

Vocabulary size

Grimaldi is very careful when discussing vocabulary size in CL: "I say that there are not, probably, more than six hundred primitive words, because I have not, after years of search, discovered more than that number, and am of the opinion that the spoken words will not number more" (pp. 103–104). However, Grimaldi reminds us that Signs constitute the central means of conveying messages in CL, and that spoken words "are never used excepting when actual necessity requires their use" (p. 104). Be that as it may, 600 words can be regarded as a sizeable word list, especially for an animal that resorts to spoken language only when "signs would fail" (p. 104).

Words in common use

Grimaldi (p. 114) provides a list of seventeen of the most important words in CL, which includes "Aeilo" (food), "Lae" (milk), "Aliloo" (water), "Bl" (meat), "Ptlee-bl" (mouse meat), "Bleeme-bl" (cooked meat; an unexpected item in CL), "Pad" (foot) and, of course "Mieouw" (here).

Number system and time expressions

Grimaldi is (and was!) the first to express his surprise when realizing how rich CL is when it comes to time expressions, which is partly based on the extraordinarily rich number system CL possesses.

Let's start with the latter. The CL number system seems to be a base-ten system, where the figures 1–12 have unique names, 13–19 have an ending "-do(o)" (tantamount to the English "-teen"), and 20, 30 (etc) add the basic numbers 1–9 to the stem (exactly like English), i.e.:

1="Aim"; 2="Ki"; 3="Zah"; 4="Su"; 5="Im"; 6="Lah"; 7="El"; 8="Ic"; 9="No"; 10="End"; 11="Est"; 12="Ro"; 13="Zah-do"; 14="Sudoo" (etc); 20="Ki-le"; 21="kile-aim" (etc); 30="Zah-le"; 40="Su-le"; 50="Im-le" (etc); 100="Aim-hoo"; 1000="milli" and "zule" means "millions" (pp. 109–110).

Most surprising of it all, however, is perhaps the fact that there is also a word for millionaire: "zuluaim". That cats can abstract from numbers to concepts like "millionaire" is, of course, nothing short of astonishing.

Turning to expressing the time of day, CL simply employs its rich number system to create expressions like: "ro sule-im" for 12:45, "im imle-im" for 5:55 etc (p. 111).

Word order

Grimaldi: "In the feline language the rule is to place the noun or the verb first in the sentence, thus preparing the mind of the hearer for what is to follow (p. 116). Example sentences (which one might assume are all authentic) are: "Milk give me", "Meat I want", "Sick I am", "Going out, my mistress" and "Happy are my babies" (p. 116).

Here it would, perhaps, be of interest to learn what the word order of Dog Language, or Horse Language (etc) might be, in order to create a mammal language word order taxonomy?

Prosodic inflection

When translating the words of CL it must "constantly be kept in mind" (p. 117) that prosodic inflection is central in conveying specific meanings of words. Grimaldi lists the following, elucidating, examples:

"Meouw", produced with:

1. Ordinary tone, means "how", "Good morning" or "How d'ye do?" etc;
2. Strong emphasis and a high tone on the first syllable "me", means "hatred";

"Purrieu", produced with:

1. A long roll of the "r" and a rising inflection indicates a mother calling her kittens;
2. A shrill inflection to the last syllable is "a note of warning to her loved ones" (p. 117);

"Yew", produced:

1. As an explosive, is the cat's strongest expression of hatred, and is a declaration of war (p. 117);
2. In an ordinary tone, indicates that the speaker is not feeling well.

"Poopoo", produced with:

1. Slight emphasis on the first syllable, means "sleep";
2. Strong emphasis on the last syllable means "work".

Summing up, Grimaldi points out that "there is scarcely a word in the feline language whose meaning is not subject to four or more directly opposite interpretations, according to the inflections given in its expression" (p. 117).

Grimaldi acknowledges the fact that prosodic inflection is used in human language, too: "instance the Chinese in particular. The number of words in their language is not great, but in speaking they vary each of their words by not less than five different tones, by which they make the same word signify five different things" (p. 54). The striking similarity between CL and Chinese does not go unnoticed to Grimaldi: "The Chinese language is more nearly like the Cat language than any of the existing languages, and so closely resembles it in very many respects as to almost persuade me that the language of the Cat was derived from it" (p. 104), an hypothesis strengthened by the fact that "no people are more fond of the feline than the Chinese" (p. 104).

Perception of speech and language

Turning to the perceptual capabilities of the Cat, Grimaldi refers to "Prof. William Lindsay, M. D., F. R. S., F. L. S., Hon. Member New Zealand Institute", who in his "remarkable work, entitled 'Mind in the Lower Animals' /…/ asserts that Cats readily comprehend and thoroughly understand man's words and the conversation of men" (p. 66).

The Importance of Signs

Finally, it must be pointed out that impressive as it is, CL is still overwhelmingly a language of "Signs", something which cannot be done justice by a "tiresome, misleading and fallacious grammar", or "stuffy, lame, meaningless dictionary" or "hobbling treatise upon syntax" (p. 119). The 600 (or so) "words" uttered vocally are quite obviously hugely surpassed by the expressions conveyed by the language of the ear, tail, limb, body, mouth, nose, eye, brow, chin, lip, and whiskers. Some of these signs are described by Grimaldi, but since there is not enough space here to include his examples I will only list one: when the tail "inclines toward the floor it says that its mistress may go shopping without an umbrella" (p. 121).

Concluding remarks

Grimaldi ends his paper by pointing out all the traps of human languages (mainly comparing metaphors in English and French), obviously all inferior to CL. He finishes by hoping "for better things for my favourite, the Cat" (p. 122) and bids the reader au revoir.

On the final page of his œuvre Marvin Clark expresses the expectation (and hope, one might presume) that he would "hear from" (p. 123) Grimaldi again. The story does not reveal whether this actually transpired, but Clark nevertheless expresses his firm conviction that there "can be no doubt that with the aid of the phonograph and other modern instruments /… / great progress will be made in translating and disseminating the feline language" (p. 123). And this, Clark establishes, "is a subject of vast importance" (p. 123).

Given the recent interest in the field, notably by Dr. Schötz, perhaps a corroboration of Grimaldi's findings will be made, and Dr. Schötz might in the future build upon and refine the observations made by Grimaldi and come up with an updated vocabulary list and so on and so forth. (Although I'll hedge my bets on that one.)

In any case and summing up, Grimaldi's devotion to the subject matter cannot be doubted, and this in and by itself might perhaps serve as a source of inspiration for future decoders of the (fascinating) feline language.

Addendum

Please note that the views expressed by the author in this paper do not necessarily represent the views held by the author of this paper.

Acknowledgements

I would like to thank Susanne Schötz for making me aware of the Clark/Grimaldi publication, and also for her devoted work along the same lines.

References

Champfleury (1869). *Les chats. Histoire – mœurs – observations – anecdotes.* Paris. J. Rothschild.

Clark M (1895). *PUSSY and Her Language. Including a Paper on the Wonderful Discovery of the Cat Language by Alphonse Leon Grimaldi, F. R. S., etc.* [Publisher unknown.]

Grown J (2014). *All You Need To Know About Cats*. Neil Playfoot Publisher.

Moelk M (1944). Vocalizing in the House-Cat: A Phonetic and Functional Study. *The American Journal of Psychology* 57(2): 184–205.

Schötz S (2013). A phonetic pilot study of chirp, chatter, tweet and tweedle in three domestic cats. In: Eklund R (ed.): *Proceedings of Fonetik 2013,* 12–13 June 2013, Linköping University, Sweden, 65–68.

Languages with pulmonic ingressive speech: updating and adding to the list

Robert Eklund
Department of Culture and Communication, Division of Language and Culture, Linköping University

Abstract
Speaking on inhalation, pulmonic ingressive speech, is well-known in Scandinavia and often believed to be unique to this part of the world. It has, however been shown (Eklund, 2002, 2007, 2008) that not only is ingressive speech not confined to the northernmost part of Europe, it is found all over the world and might be regarded as a linguistic universal, and can be placed in one of the different universal categories described by Croft (2003). In connection with the Eklund (2008) publication, a website was created, devoted to ingressive speech and phonation: http://ingressivespeech.info. Over the years incoming comments and reports have both offered further evidence for languages already on the list, as well as new languages with ingressives. Some of these are described in this paper.

Introduction

In the beginning of 2015, journalist Oliver Gee, working for the magazine *The Local* (http://www.thelocal.se), published a clip on YouTube that went viral (Gee, 2015). It made headlines in Daily Mirror (16 January 2015), with the title "Bizarre noise for 'yes' word in Swedish language will blow your mind", Dagens Industri (22 January 2015) with the title "Norrländskt ord fascinerar världen", and even in Australia, where news.com.au (16 January 2015) reported that in Swedish, instead "of a word, a quick intake of breath through pursed lips indicates the affirmative". Other sources that covered the Gee's interview include Huffington Post, and (according to Dagens Industri) even Fox News.

That the Swedish habit of speaking on inhalation is known, and regarded as somewhat strange, is nothing new. In fact, there are sources galore that cover this phenomenon, and that the "jo" (affirmative 'yes') perhaps is especially common in northern Sweden has also been discussed (e.g. Salö, 2007). However, that ingressive speech is unique to Sweden has been proven to be not exactly true, even if ingressive speech probably is an unusually frequent phenomenon in northern Europe (Eklund, 2002, 2007, 2008).

Of special interest here is that not only has the "uniqueness myth" been "debunked", it has been discussed in The Local, where journalist Salomon Rogberg previously has interviewed Robert Eklund about this allegedly Swedish phenomenon, which resulted in an article with the title "Swedes and donkeys: a language peculiarity" (The Local, 18 December 2012), which despites its title actually points out that ingressive speech occurs all over the world, e.g. in Canada, the Philippines and Greece, and establishes that "So *is* the northern vacuum cleaner unique to Sweden? Probably not." (It seems Oliver Gee was not aware of Salomon Rogberg's article; Personal Communication, 23 January 2015).

However, seeing how ingressive speech still can make headlines across the globe, it seemed in place to provide an "update" as to what languages employ ingressive speech, based on correspondence received through the website http://ingressivespeech.info over the past years.

An ingressive primer

To make a long story very short, ingressive speech has been around for thousands of years, and it was once thought that ventriloquists used ingressive speech as their "trick", with an early mention already in 1657 by van Helmont (1657:22). The first mention of ingressive speech as a paralinguistic phenomenon, very much used the way it is used today (as a version of 'yes') is mentioned already in 1765 by Cranz (1765:279) when discussing the language of the Eskimo. Eklund (2008) lists around 50–60 languages where ingressive speech has been reported in the literature, although not all sources mention exactly *what* language(s) in a specific region make(s) use of ingressive speech.

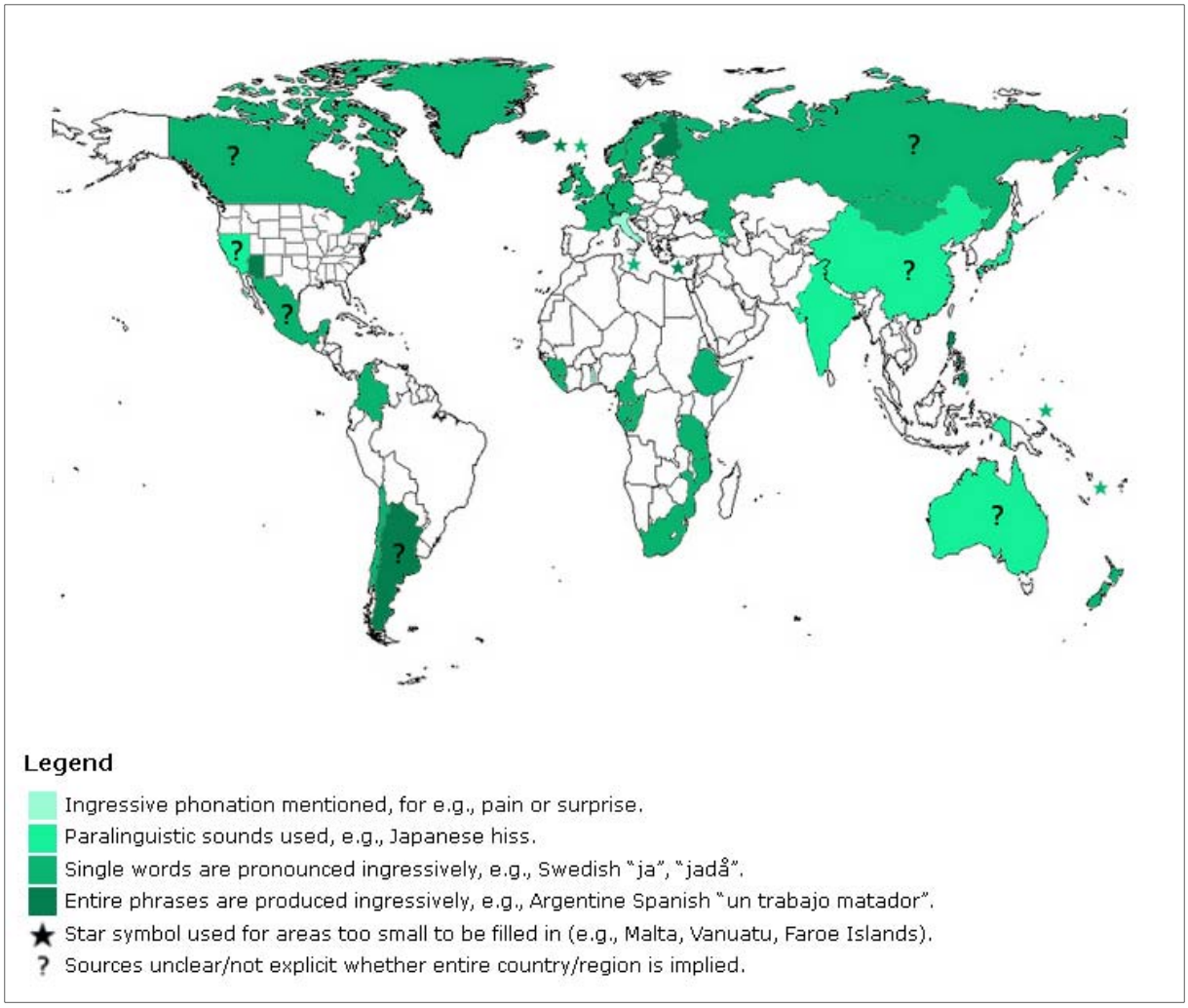

Figure 1. Ingressive speech map retrieved from http://ingressivespeech.info on 2 May 2015.

The most interesting point, however, is that those languages are found all over the world, in languages belonging to very different language groups, so although it seems safe to establish that while the *frequency* of ingressive phonation surely is very elevated in northern Europe (e.g. the Baltic states, Finland, Scandinavia, Iceland, the Faroe Islands), the *occurrence* of ingressive speech beyond doubt is global.

In connection with Eklund (2008) a webpage was created to accompany the JIPA paper (Eklund, 2008), but the website also covers other ingressive phonation types (like cheetah purring) and provides an updated language map where ingressive speech phenomena are displayed in a graded fashion, using different colors and a legend.

This map, as it appeared on 2 May 2015, is shown in Figure 1 above.

An updated list

Over the past years, ingressivespeech.info has resulted in several emails from interested people both enquiring about ingressive speech in general, but also reporting its use in languages either already covered/mentioned in Eklund (2008) or not mentioned elsewhere.

In the following sections I will summarize some of these, either lending support languages that are already covered, or adding new languages to the list. The focus will be on "substantiated" and first-hand mentions.

Known languages/countries

Below I will list reports received on languages or regions that are already known to exhibit ingressive speech.

Canada

Timothy Cummings (P.C., 28 May 2011) confirms ingressive speech in Canada, on Prince Edward Island, and in Moncton, where at least one local is known to "ingress her 'yes's".

New Zealand (English)

Timothy Cummings (P.C., 28 May 2011) reports that when he lived in New Zealand during the period 1997–2002, he heard "a handful of Kiwis ingress their 'yes's." He also reports that this mostly occurred in the village of Havelock North on the North Island.

Khalkha Mongolian

Colton Wiscombe, an student in theoretical linguistics at the University of Utah, lived in Mongolia during the period 2006–2008 and confirms (P.C:, 29 October 2009) that there are several words in Khalkha Mongolian that can be produced ingressively, including (but not limited to) тийм ("that"/ "yes"), үгүй ("no") and мэдэхгүй ("I don't know").

New languages

In this section I list comments that describe ingressive phonation in languages that were not covered in Eklund (2008), and seem to be lacking from the academic literature.

Ethiopia and Amharic

Sharon Cottrell (P.C., 25 March 2009) reports that "ao" (meaning "yes") was produced ingressively while she was working in Ethiopia during the period 1973–1975. When she later visited Denmark, she was "amused to find that Danes used the same sound".

Timothy Cummings (P.C., 28 May 2011) tells me that several Ethiopians he met in 1996 "inhaled while saying the Amharic equivalent of "yes": 'ow'" (compare "ao" above), and that they "inhaled very noticeably".

Ryan Johnson (P.C., 1 November 2013) reports that a waitress in an Ethiopian restaurant in Minneapolis uttered longer phrases like "you're welcome" and "thank you" ingressively. When asked what her native language/s was/were, she mentioned Amharic, English and Swahili. Johnson was intrigued by this, given that he speaks Finnish and was consequently aware of the phenomenon, since, as he says that Finnish "allows for phrases and sentences to be inhaled without anyone really batting an eyelash".

Further support for the occurrence of ingressive speech and phonation in Ethiopia (in both Amharic and English) comes from Robert Fultz (P.C., 16 July 2014) who informs me that while visiting Addis Ababa, he noticed that "pulmonic ingressive" phonation was used "as a kind of agreement/acknowledgement while another person was speaking", i.e. as a back channel. Fultz further reports that speakers used ingressive phonation on the items "yeah" and "right", while nodding at the same time, and that all the speakers he heard using ingressives were "male, educated, and probably 25–45".

Albanian

Klaus E. Gjika (P.C., 7 April 2015) reports that there are two words for 'yes' in Albanian, "po" and "e" (the former seems to be the "official" one), where the latter is sometimes produced ingressively. Moreover, as Gjika writes: "there is also another affirmative, which sounds like an ingressive / ɸ /, placing your lips and tongue in a whistling position, but sucking air instead of blowing. To my understanding, this isn't very dissimilar to how Northern Swedish dialects produce their ingressive affirmation".

Interestingly, Gjika also reports that: "It's noteworthy though that younger people from 30 years old and below don't have ingressive sounds, as far as I can tell at least. Or maybe they don't in the city I live in, Vlorë." This is similar to what seems to be the situation in other parts of the world, e.g. Newfoundland, where ingressive speech and phonation seems to (more or less) be a thing of the past, and where recent reports, or reports of young people phonating ingressively are rare or non-existent.

Final comments

Comments on universality

As has previously been pointed out, ingressive speech is not a linguistic phenomenon which is unique to Sweden (or Norway), and Eklund (2008:283–284) assessed to what degree ingressive speech actually instead can be regarded as a universal, following definition 2 in Croft's listing of different kinds of universals (2003:236), i.e.:

1. Linguistic phenomena that are areally widespread, and common in genetically closely related languages may be frequent and stable. Examples include front unrounded vowels.

2. Phenomena that are widespread but relatively sporadic within genetic groups may be frequent but unstable; examples include nasal vowels and definite articles.
3. Phenomena that are relatively scarce in the world's languages, but common in genetic groups in which they occur, may be stable and infrequent. Examples include vowel harmony and verb-initial word order.
4. Finally, phenomena that are both scarce and sporadic may be unstable and infrequent; examples include velar implosives and object-initial word order.

It would seem that the more we look at ingressive speech, the less "highly marked" (Clarke & Melchers, 2005:51) it seems to be.

Source reliability

It is obviously difficult to assess the reliability of the sources cited in this paper. Email correspondence from "unknown" sources without any "hard data" to back up the claims, is, of course, not the ideal foundation for any claims of scientific strength. However, the data sources referred to in this paper are not all that different from many sources cited in Eklund (2008), where several mentions are "anecdotal", and even more are not in any way supported by sound files or other reliable material.

For example, Key's (published) comment that in Scandinavian, "the women but not the men, express agreement by articulating 'ja' with air drawn in" (Key, 1975, 150) is obviously incorrect. However, one can use this quote in order to illustrate how this field is replete with dubious sources, and is also characterized by urban myths, much like the viral video which was mentioned in the introduction. Having corresponded with the informants mentioned in the present paper gives me no reason to be more skeptical about their reliability than many of the sources referred to in my previous publications.

Acknowledgments

I would like to thank everyone who has submitted comments to the website, whether these comments have been included in this paper or not. Several people have taken a substantial amount of both time and effort to answer my questions, and some of the descriptions have been phonetically very impressive, especially considering that most comments come from non-phoneticians. So, not listing everyone, but with no one forgotten, I would like to extend my thanks to the following: Sharon Cotterell, Alan Clifford, Timothy Cummings, Klaus E. Gjika, Robert Fultz, Ryan Johnson and Colton Wiscombe.

References

Clarke, S & Melchers, M (2005). Ingressive particles across borders: Gender and discourse parallels across the North Atlantic. In: M Filppula, J Klemola, M Palander & E Penttilä (eds.), *Dialects across borders: 11th International Conference on Methods in Dialectology* (*Methods XI*), *Joensuu, August 2002* (Current Issues in Linguistic Theory 273), Amsterdam: John Benjamins, 51–72.

Croft, W (2003). *Typology and universals* Cambridge: Cambridge University Press.

Dagens Industri (2015). *Norrländskt ord fascinerar världen*, 22 January 2015. www.di.se/artiklar/2015/1/22/norrlandskt-ord-fascinerar-varlden/

Daily Mirror (2015). *Bizarre noise for 'yes' word in Swedish language will blow your mind*, 16 January 2015 http://www.mirror.co.uk/news/weird-news/bizarre-noise-yes-word-swedish-4988454

Eklund, R (2002). Ingressive speech as an indication that humans are talking to humans (and not to machines). *International Conference on Spoken Language Processing (ICSLP) 2002*, Denver 2:837–840.

Eklund, R (2007). Pulmonic ingressive speech: A neglected universal? *Fonetik 2007*, Stockholm, *TMH-QPSR, KTH*, vol. 46, 21–24.

Eklund R (2008) Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association* 38(3):235–324.

Gee, O (2015). *Is this the most unusual sound in the Swedish language?* www.youtube.com/watch?v=URgdIAz4QNg

Key, M R (1975). *Paralanguage and kinesics*. Metuchen, NJ: The Scarecrow Press.

News.com.au (2015). *People in Northern Sweden have the world's weirdest way of saying 'yes'*, 16 January 2015. http://www.news.com.au/lifestyle/real-life/people-in-northern-sweden-have-the-worlds-weirdest-way-of-saying-yes/story-fnq2oad4-1227187456841

Rogberg, S (2015). Swedes and donkeys: a language peculiarity. *The Local*, 18 December 2015. http://www.thelocal.se/20121218/45124

Salö, L J (2007) *.jo – en studie i användning av ett ingressivt talfenomen*. Ms., term paper, Umeå University.

A spectral analysis of the backing of Afrikaans /s/ in the consonant cluster /rs/

Otto Ewald
Center for Languages and Literature, Lund University, Sweden

Abstract

This acoustic study investigates the phoneme /s/ in Afrikaans and its backed realization in the consonant cluster /rs/ in coda position. The study focuses on spectral differences between the two contexts, with and without a preceding /r/, employing two spectral measurements: the center of gravity and skewness.

The analysis revealed a lower center of gravity and a lesser degree of skewness for /s/ in the cluster /rs/, indicating a flatter spectrum with more energy at lower frequencies, pointing at more back place of articulation and the realization of /s/ as a voiceless retracted alveolar sibilant [s̠] in this context. Future studies should aim at investigating additional acoustic properties of /s/ as well as /r/ in different contexts in Afrikaans to describe their exact realizations and their interaction in further detail.

Introduction

Sibilants feature a wide range of possible places of articulations and can further be described according to which part of the tongue comes in contact with the passive articulator during their production (Ladefoged and Maddieson, 1996). This gives rise to a large number of possible sibilant articulations, and earlier research has shown that the exact production of a sibilant phoneme (Dart, 1991) as well as its spectral characteristics (Hughes and Halle, 1956; Lindblad, 1980) can vary between speakers of the same language, although this might not be the case for all languages (Gordon et al., 2002). Furthermore, it has also been suggested that there is a tendency for greater articulatory variability in languages with a smaller set of sibilant phonemes such as English and French, than in those with a larger set, such as Mandarin Chinese and Swedish (Toda and Honda, 2006).

Acoustically, there are several cues to a sibilants place of articulation. As a general rule, the more front a sibilant's place of articulation is, the higher will the cut-off frequency be between low amplitude frequencies and high amplitude frequencies, which correlates with a decrease in size of the frontal cavity (Ladefoged and Maddieson, 1996; Lindblad, 1980).

Another acoustic measurement is the location of the frequency peak (Adam, 2012) or spectral peak frequency (Cheon and Anderson, 2008; Newman, 2003; Jongman et al., 2000), which has been found to be a good differentiator between for example /s/ and /ʃ/ in English.

Measurements that describe the whole spectrum have also been used to describe sibilants, such as the center of gravity and skewness. The former is a measure of the average frequency of the spectrum, while the latter indicates the asymmetry in the distribution of spectral energy in the frequency domain (Jongman et al., 2000).

Afrikaans, a West Germanic language spoken in South Africa, features an excellent opportunity to study sibilant allophony. /s/ is the only native phoneme, with /ʃ/ occurring as a foreign phoneme in a handful of loanwords such as *sjampanje* /ʃamˈpanjə/ 'champagne', *chirurg* /ʃiˈrœrx/, 'surgeon' (Donaldson, 1993), and *masjien* /maˈʃin/ 'machine'. Following Toda and Honda (2006), a larger allophonic variation in /s/ can be expected since the sibilant inventory is small, and Canepari and Cerini (2013) describe a range of phonetic realizations of /s/: dental, dentoalveolar, and alveolar. They note that alveolar realization is common in the consonant clusters /rs/, /sl/, /st/, and /sk/.

The backing of /s/ in the cluster /rs/ is deemed to be of particular interest here due to its similarity to the supradental realization of the cluster /rs/ as [s̪] in western, northern, and central Swedish dialects (Bruce, 2010). This study explores Afrikaans /s/ in this particular rhotic context by comparing it to single /s/ in coda position, with the goal of identifying a backed allophone of /s/ in the /rs/ cluster and quantify-

ing the acoustic differences in /s/ between the two contexts using the spectral measurements of the center of gravity and skewness. The backed allophone of /s/ after /r/ is expected to show more spectral energy at lower frequencies.

Experiment

Speakers

Three native speakers of Afrikaans were recorded, one male (M1) and two females (F1 and F2). They were 19-20 years old and currently studying in Bloemfontein at the time of the recording, but are originally from Vanderkloof (Northern Cape province), Bloemfontein (Free State province), and Schweizer-Reneke (North West province).

Material

The material consisted of a list of 20 words created specifically for this study and read in the carrier sentence *Ek het X geskryf* 'I wrote X'. The target words contained a coda /s/ in either a non-rhotic context (V_) or a rhotic context (Vr_) in the stressed syllable. The nucleus of the stressed syllable was one of the five vowels /a/, /ɛ/, /ɔ/, /i/, /u/ to minimize coarticulatory effects on the target phoneme. The test words are shown in Table 1. Each word occurred twice in the experiment, yielding a total of 40 words per speaker.

Table 1. Target words arranged according to vowel context and rhotic context.

	V_	Vr_
/a/	*vas* /ˈfas/	*vars* /ˈfars/
	kas /ˈkas/	*Mars* /ˈmars/
/ɛ/	*res* /ˈrɛs/	*kers* /ˈkɛrs/
	ses /ˈsɛs/	*pers* /ˈpɛrs/
/ɔ/	*los* /ˈlɔs/	*wors* /ˈvɔrs/
	kos /ˈkɔs/	*dors* /ˈdɔrs/
/i/	*kies* /ˈkis/	*passasiers* /pasaˈsirs/
	nies /ˈnis/	*Iers* /ˈirs/
/u/	*moes* /ˈmus/	*jaloers* /jaˈlurs/
	bloes /ˈblus/	*broers* /ˈbrurs/

Recordings

The three recordings were made in Audacity with a Triton Pro+ 5.1 True Surround Headset at a sampling rate of 44.1 kHz, and saved as 16 bit wav files. The male speaker was given instructions for the recording as well as the list with the 40 test sentences. He recorded himself with guidance from the author over Skype. He then assisted in carrying out the two subsequent recordings of the two female speakers F1 and F2.

Analysis

Analysis was carried out in Praat (Boersma and Weenink, 2013), and was based on FFT spectra running from 0 Hz to 10 kHz calculated from a 40 ms time window around the midpoint of the frication noise of each instance of /s/ to ensure maximal stability and representativeness of the fricative. The center of gravity and skewness of each FFT spectrum were calculated, and t-tests were carried out to find statistically significant differences in these two measurements between /s/ in the two contexts. Each FFT spectrum was also converted into a LTAS with a bandwidth of 100 Hz, and then averaged for each speaker and rhotic context for visual analysis.

Results

Spectra

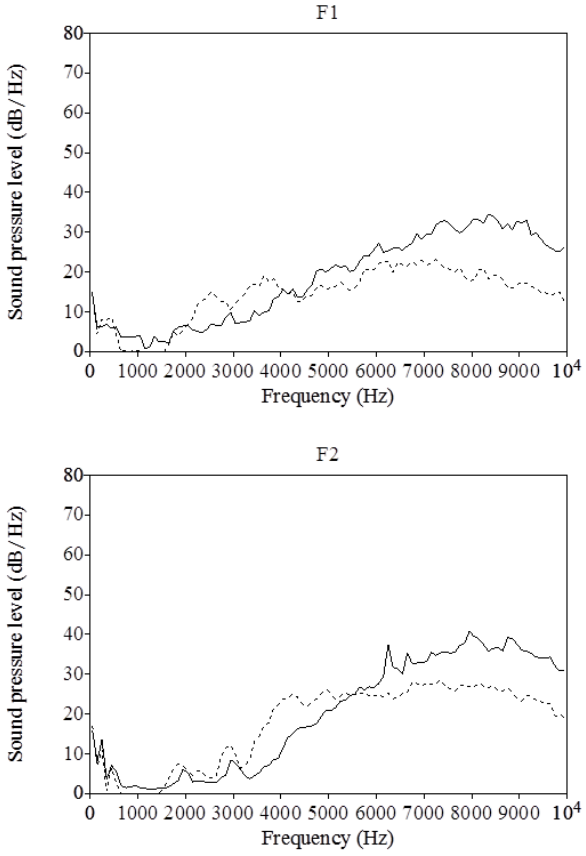

Figure 1. Average spectra for /s/ in the contexts V_ and Vr_ for F1 and F2. The V_ context is represented by a solid line and the Vr_ context by a dotted line.

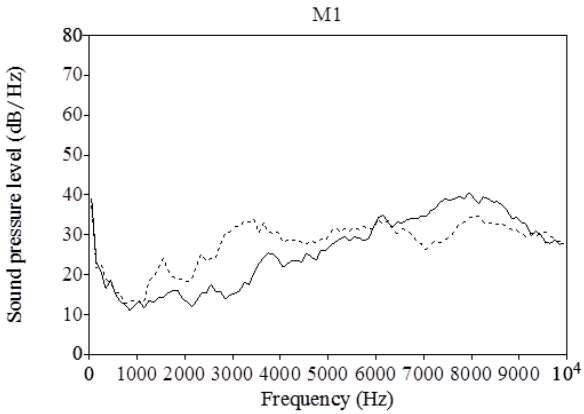

Figure 2. Average spectra for /s/ in the contexts V_ and Vr_ for M1. The V_ context is represented by a solid line and the Vr_ context by a dotted line.

The average spectra reveal a universal lowering in amplitude of higher frequencies in the Vr_ context compared to the V_ context for all speakers, as seen in Figure 1 and Figure 2. The frequency point above which this lowering takes place differs between speakers. The lowering takes place above 4400 Hz for F1, above 5500 Hz for F2, and between 6000 Hz and 9250 Hz for M1.

Each speaker also features a more varied amplification of some lower frequencies. F1 has two amplified spectral peaks centered around 2500 Hz and 3700 Hz respectively. F2 features a main amplification of frequencies between 3200 Hz and 5500 Hz, as well as two amplified peaks just below 2000 Hz and 3000 Hz. M1 has a broader band of frequency amplification between 1000 Hz and 6000 Hz.

F1 and F2 also display an additional lowering of low frequencies between 500 Hz and 1500 Hz.

Center of gravity

Table 2. The means and standard deviations for the center of gravity in the two contexts for each speaker.

	Mean (Hz)		Standard deviation (Hz)	
	V_	Vr_	V_	Vr_
F1	7847	6289	193	511
F2	8014	6852	424	440
M1	7199	5817	290	587

The center of gravity shows a considerable difference between the two contexts, as show by the mean values in Table 2 and the distribution of the data in Figure 3. /s/ in the Vr_ context features a more than 1100 Hz lower average than /s/ in the V_ context for all speakers, indicating more spectral energy at lower frequencies in the Vr_ context.

There is also a considerably larger variation in the Vr_ context, as show by the standard deviation in Table 2.

The difference in the center of gravity between the two contexts is highly significant for all speakers ($p<0.01$).

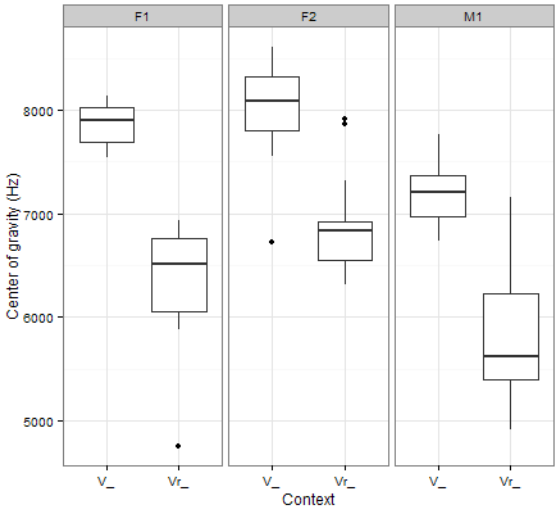

Figure 3. Box plot of the center of gravity in the two contexts for each speaker.

Skewness

Table 3. The means and standard deviations for skewness in the two contexts for each speaker.

	Mean		Standard deviation	
	V_	Vr_	V_	Vr_
F1	-1.2	-0.7	0.4	0.4
F2	-0.9	-0.6	0.8	0.4
M1	-2.5	-0.6	0.7	0.3

The differences in skewness are less clear compared to the center of gravity. Both contexts show overall negative skewness for all speakers, as can be seen in Figure 4. Although they all feature a lesser degree of skewness in the Vr_ context, the difference between the contexts is much smaller in F1 and F2 than in M1, as is evident in Table 3 and in Figure 4. The difference between the two contexts is here indicative of greater symmetry between the lower and

higher frequencies of the spectrum, as Figure 1 and Figure 2.

As for standard deviation, F2 and M1 pattern together in having larger standard deviation in the V_ context, while there is no difference between the two contexts for F1 in this case.

However, while the difference in skewness is highly significant for F1 and M1 ($p<0.01$), it is not for F2 ($p=0.14$).

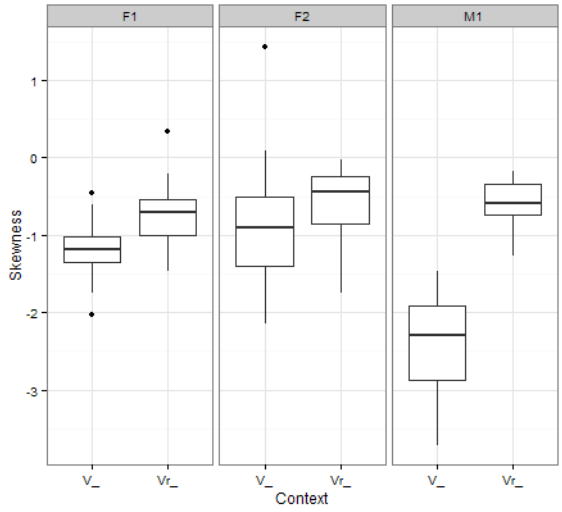

Figure 4. Box plot of skewness in the two contexts for each speaker.

Discussion

The backing of /s/

The results of the spectral analysis of /s/ reveal highly significant differences in the center of gravity for all three speakers as well as highly significant differences in skewness for F1 and M1 between the two contexts. These findings indicate a lowering in amplitude of high frequencies and more spectral energy at lower frequencies as well as a more equal distribution of spectral energy in /s/ when it is preceded by /r/, and it is possible to conclude that /s/ displays several acoustic signs of having a more back articulation in this context, strengthening Canepari and Cerini's (2013) observation and confirming the hypothesis.

The results also show parallels with Gordon et al. (2002) in that the center of gravity is a good measure of the place of articulation of a sibilant, while the measure of skewness on the other hand only gave statistically significant results for two of the speakers. This with the findings of Jongman et al. (2000), who consider skewness to be the best indicator of the place of articulation.

While it is clear that /s/ has a distinct allophone following /r/ with a more back place of articulation, it is difficult to precisely describe the articulatory setting from the results presented here, not only for /s/ in the Vr_ context but also in the V_ context. Based on Canepari and Cerini's (2013) description and the present results, the backed allophone of /s/ is tentatively labeled as a voiceless alveolar retracted sibilant [s̠].

Implications for further research

The results of the present study give rise to several new research questions to be answered in future research.

A more in-depth spectral analysis of /s/ in the two contexts should include other spectral measures such as kurtosis or could involve the measurement of the spectral peak in order to further describe the acoustic properties of this phoneme and its allophonic variation.

To increase our understanding of the interaction between /r/ and /s/ in the consonant cluster /rs/ in Afrikaans, it is imperative to also investigate the acoustic characteristics of /r/ in the V_ and Vr_ contexts, as well as other contexts, such as word-initially, word-medially, word-finally, and before other consonants such as plosives. While /r/ was not addressed in the present study, it seems to be rather weak in the /rs/ cluster, to the point where it is almost inaudible. /s/ itself should also be studied in the same different environments to better describe its allophonic behavior.

Going beyond acoustic phonetics, an articulatory study using electropalatography or articulography would be immensely helpful in describing the articulatory setting during the production of the allophones of /s/ in the two environments. It is however clear that such a study would be much more difficult to carry out with regards to the kind of equipment needed and the availability of participants.

The similarity between the described allophonic behavior of /s/ after /r/ in Afrikaans and the realization of the same cluster as a supradental [s̪] in many Swedish dialects encourages cross-linguistic comparison between the sibilants in the two languages and might suggest, together with data on the behavior of /r/ in this context, that the /rs/ cluster in Afrikaans is going through the same change as its Swedish counterpart historically has done.

Conclusion

This study has explored the allophony of the /s/ phoneme in Afrikaans in two contexts, one without a preceding /r/ and one in the consonant cluster /rs/, employing the spectral measurements center of gravity and skewness. /s/ in the cluster /rs/ was found to have a lower center of gravity and a lesser degree of skewness, confirming the existence of a backed allophone of /s/ in this context, which has been tentatively described as a voiceless alveolar retracted sibilant [s̠].

Future research should aim at describing this retracted allophone of /s/ in further detail by using a greater range of acoustic measures and by studying it in different contexts. There is also a possibility for a cross-linguistic comparison between Swedish and Afrikaans when it comes to the consonant cluster /rs/ due to the similarity of its realization in the two languages, which might also be pointing at a sound change under way in Afrikaans.

References

Adam H (2012). An acoustical study of the fricative /s/ in the speech of Palestinian-speaking Broca's aphasics - Preliminary findings. *Linguistik Online,* 53.

Boersma P, Weenink D (2013). Praat: doing phonetics by computer (Version 5.3.60).

Bruce G (2010). *Vår fonetiska geografi. Om svenskans accenter, melodi, och uttal.* Lund: Studentlitteratur.

Canepari L, Cerini M (2013). *Dutch & Afrikaans pronunciation & accents.* München: LINCOM .

Cheon S, Anderson V (2008). Acoustic and perceptual similarities between English and Korean sibilants: implications for second language acquisition, *Korean Linguistics*, 14: 41-64.

Dart S N (1991). Articulatory and acoustic properties of apical and laminal Articulations. *UCLA Working Papers in Phonetics,* 79.

Donaldson B (1993). *A Grammar of Afrikaans*. Berlin-New York: Mouton de Gruyter.

Gordon M, Barthmaier P, Sands K (2002). A cross-linguistic acoustic study of fricatives. *Journal of the International Phonetic Association*, 32: 141-174.

Hughes W, Halle M (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America* 28: 303-310.

Jongman A, Wayland R, Wong S (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108 (3): 1252-63.

Ladefoged P, Maddieson J (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.

Lindblad P (1980). Svenskans sje- och tje-ljud i ett Allmänfonetiskt Perspektiv. *Travaux de l'Institut de Linguistique de Lund 16*. Lund: C. W. K. Gleerup.

Newman R S (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America,* 113: 2850-60.

Toda M, Honda K (2003). An MRI-based crosslinguistic study of sibilant fricatives. *Proceedings of 6th International Seminar on Speech Production*. Australia.

Self-perception of vocal and articulatory effort in consonant production by native Swedish speakers

Iris Gordon-Bouvier[1], *Josefine Kyhle*[1], *Anita McAllister*[1], *Hanna Norman*[1], *Sarah Paues*[1], *Camille Robieux*[2] *and Sofia Strömbergsson*[1]
[1]*Karolinska Institute, Stockholm, Sweden*
[2]*University of Marseille, France*

Abstract

This study investigates native Swedish speakers' self-perception of vocal and articulatory effort in a speech sound production task, in an attempt to identify patterns amongst Swedish speakers as well as to make cross-language comparisons with previously reported data for French speakers. Vocal and articulatory effort refers to the level of exertion required to produce a sound. Previous research has primarily focused on physiological rather than self-perceived measurement.

In a partial replication of a French study on vocal and articulatory parameters, twenty-two healthy native speakers of Swedish aged 21-50 were presented with 220 pairs of nonsense syllables that they were asked to produce aloud. Each pair included up to three contrasts between the consonants in terms of context, voicing, and place and manner of articulation. Some subjects were instructed to select the item from each pair that they felt was easier to produce, while others were instructed to choose the one they perceived as more difficult.

The results indicate that subjects perceived voiced consonants as requiring a greater level of effort than voiceless consonants. A higher rate of self-perceived effort was also reported for unvocalic (isolated) consonants compared with intervocalic consonants, regardless of voicing.

These findings may be significant for professionals in the field of speech therapy, by providing a baseline for structuring clinical voice therapy based on subjective experience.

Introduction

During speech, the glottis varies from wide open to tightly shut, due to variations in the level of air pressure produced by the lungs and laryngeal muscle activity (Johnson, 2012).

When air particles travel through a narrow passage such as the glottis, the particles accelerate and thin out, lowering air pressure in the passage. Due to the elasticity of the glottis, the low air pressure causes the vocal folds to contract and the airflow is interrupted. The passage remains closed until the subglottic pressure is built up to the point where it pushes the glottis open again, and the cycle repeats over and over. Voicing, or phonation, occurs when the vocal folds vibrate in this manner, according to the myoelastic-aerodynamic theory of voice production (Van den Berg, 1958).

Vocal fold vibration is the usual source of sound in vowels, and our vocal tract serves as an acoustic filter that modifies the sound (Johnson, 2012). This model is known as the source-filter theory of speech production (Fant, 1960).

Consonants are sounds that are produced through a narrowing or complete closure anywhere between the glottis and the lips, using the active articulators, i.e. the parts of the vocal tract that can be manipulated at will (Engstrand, 2004). Depending on whether or not the vocal folds are engaged, consonants are either voiced or voiceless.

Vocal and articulatory effort refers to the level of exertion required to produce a sound. Many patients with voice disorders, such as dysphonia, nodules or polyps, report a general increase in vocal effort (Rosenthal et al, 2014).

In a clinical situation, a speech therapist can recommend and train patients in various vocal

exercises, with thoroughly documented effects (ASHA, 2005). However, beyond patients' informal reports of improved vocal function, these effects are generally based on physiological measurements of human vocal effort rather than scientific studies of self-perception in healthy subjects. In a study, focusing on laryngeal resistance (LR, subglottic pressure divided by average airflow through the glottis) a correlation was found between LR and a trained professional's auditory perception of participants' vocal quality (Grillo et al., 2009). To date, limited research has been done into the physiological parameters that contribute to the perception of vocal and articulatory effort. The present study aimed to understand how Swedish healthy subjects perceive vocal and articulatory effort in a production task, and whether the findings differ from those made in a preliminary study performed on French speaking subjects (Robieux, 2015). Robieux's preliminary results indicate that native speakers of French perceive that a significantly greater effort is needed to produce voiced consonants than voiceless regardless of context. The two main research questions of the present study were whether there would be corresponding patterns in the self-perception of Swedish speech sound production, and to investigate differences and similarities between Swedish and French results.

There are a number of articulatory differences between Swedish and French. One of these is aspiration. In Swedish, voiceless plosives are usually aspirated when they occur in an initial position within a stressed syllable, unless preceded by /s/ (Engstrand, 1999). In verbal communication prior to the study, Robieux expressed an expectation that aspiration on voiceless plosives could affect perceived vocal and articulatory effort to the extent that Swedish subjects might rate voiceless consonants as more difficult to produce than voiced. As native speakers of Swedish, the authors of the present study proposed a somewhat different hypothesis, namely that the differences between the two languages would not be enough to override the general increase in effort required for phonation. Voiced consonants involve a high level of coordination between subglottic pressure and the vocal folds, together with a constriction or complete closure of the vocal tract (Ohala, 1983). Furthermore, these changes occur extremely fast, in a matter of milliseconds. The expectation was that the results would reflect this increased level of activity. A secondary hypothesis was that unvocalic (isolated) consonants would be perceived as more effortful than intervocalic consonants (uttered between two vowels), due to the need to coordinate high air pressure, vocal articulators and, in the case of voiced consonants, phonation, in a very short space of time.

The present study may provide valuable insight that can be exploited for the purpose of improving voice therapy. Identification of those speech sounds perceived by healthy subjects as requiring more effort to produce can help caregivers to adapt treatment by reducing the sense of vocal effort while still effectively treating voice disorders.

Method

Subjects

A convenience sampling of twenty-two adult subjects was carried out, with thirteen women and nine men ranging in age from 21-50 years. All subjects were native speakers of Swedish, with no history of reading, speech or voice disorders.

Stimuli

Each subject was presented with 220 pairs of items made up of nine different consonants, both voiced and voiceless - /b/, /p/, /d/, /t/, /g/, /k/, /v/, /f/, and /s/ - combined with the vowel /a/. The consonants contrasted in terms of place and manner of articulation, and comprised labial, dental and velar plosives and fricatives. Consonants were presented in four different contexts: nonsyllabic (unvocalic #C#), monosyllabic (prevocalic #CV, postvocalic VC#) and disyllabic (intervocalic VCV). Pairs were presented randomly, each pair occurring twice in reverse order; item 1-item 2 versus item 2-item 1. There were two different test forms; one instructing subjects to focus on the sounds that felt easier to produce, and the other instructing subjects to select the more difficult option. Both tests contained the same pairs of items, but in a different order.

Procedure

The experiment was of a between-subjects design, and was carried out in a quiet, familiar setting. Subjects were given both verbal and written instructions, and were asked to sign a

consent form stating that they agreed to their data being used in the study.

Prior to starting the test, participants were presented with a list of speech sounds for training, enabling experiment leaders to correct pronunciation if necessary, for example if a voiceless consonant was pronounced in a voiced manner. Such corrections were also made during the actual test.

The task consisted of reading each pair aloud using a normal speaking voice, with a brief pause between the two items, and to circle one item per pair, depending on specific instructions. Repetition of each item was permitted as necessary, and participants were instructed to leave no blanks. Answers were recorded by hand, and no digital recordings were made. The test was completed in a single session, with a mean time of approximately 5-8 minutes per page, although there was no time limit.

Data were entered manually in an Excel spreadsheet, with one data sheet per subject, and answers were converted to the dichotomous values 1 and 2. χ^2 analyses were then performed on the relevant parameters using the SPSS software program.

Results

Two primary contrasts were selected and analysed; voiced-voiceless and unvocalic-intervocalic.

Combined responses from all participants rendered 704 instances where the main contrast was whether consonants were voiced or voiceless. In 407 (57.8 %) of these cases, subjects chose the voiced consonant as requiring more vocal effort (see *figure 1).*

There were 396 total instances where the main contrast between the consonants was unvocalic (#C#) versus intervocalic (VCV). In 259 instances (65.4 %), the unvocalic consonant was selected as requiring more effort to produce *(*see *figure 2)*. A χ^2 analysis of these primary contrasts yielded a significant result ($p<0.001$) for both parameters. For the contrast voiced-voiceless χ^2 (1, N=704) = 17.19, $p<0.0001$. For the contrast unvocalic-intervocalic χ^2(1, N=396) = 37.7, $p<0.0001$.

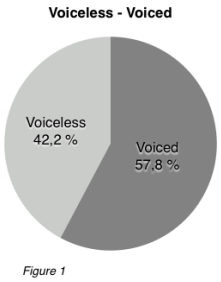

Figure 1

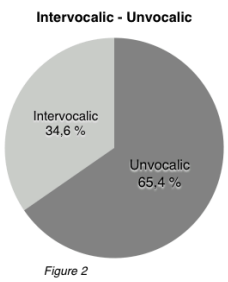

Figure 2

Figure 1 shows how test subjects rated vocal effort required to produce consonants where the main contrast was voiced-voiceless, while the main contrast in figure 2 was unvocalic-intervocalic. Both figures show how often subjects rated the parameters as requiring more effort.

Discussion

The two questions that were posed within the present study were whether a significant pattern would be found in Swedish speakers' self-perceived vocal and articulatory effort in producing consonants, and how this would compare to data from the French study (Robieux, 2015).

The main hypothesis underlying this study was that Swedish speakers would rate voiced consonants as being more difficult to produce than voiceless consonants, hence that our result would match those rendered by Robieux. A secondary hypothesis was proposed, namely that unvocalic (isolated) consonants would be perceived as requiring more vocal effort than intervocalic consonants.

Indeed, analysis of data yielded a significant result, in alignment with both the primary and secondary hypotheses. In the majority of cases where subjects were asked to choose between voiced and voiceless consonants, voiced were rated as being more difficult than unvoiced, corroborating Robieux's findings. In the majority of the cases where the main contrast was unvocalic-intervocalic, unvocalic consonants were deemed harder to produce.

Interpretation of data

The production of a plosive requires a complete closure of the vocal tract, resulting in a build-up of pressure, which is subsequently released when producing the sound. If the plosive is voiced, the vocal folds are activated before or upon the burst. This subglottal pressure build-up and release may require more effort, also involving the coordination of the speaker's

vocal folds and vocal tract when the plosive is voiced (/b/, /d/, /g/). If the plosive is voiceless, (/p/, /t/, /k/), the effort may be smaller.

When producing an open vowel, pressure from the lungs is constant, creating a flow through the vibrating vocal folds. If a plosive is produced in an intervocalic context, the production of the plosive may be perceived by the speaker as "riding the wave" of this subglottic pressure thus appearing to require less effort. Furthermore, there is some evidence that in a VCV context, where the consonant is a plosive, the movement of the articulators occur almost simultaneously (Gay, 1977). This could also explain why subjects rated intervocalic consonants as requiring less vocal effort.

It could be argued that isolated consonants rarely occur in spontaneous speech in Swedish. Therefore they may be less automated than sequences combining them with vowels would be. However, in traditional voice therapy exercises using isolated consonants are not uncommon (Carlsson et al., 1985).

Possible factors affecting outcomes

The test material was originally designed for French speakers. No instruction was given to the Swedish speakers with regards to pronunciation or duration of the open vowel sound. In Swedish, the vowel sounds /a/ and /ɑ/ are both possible, a fact that may have affected the results (Handbook of IPA, 1999). The rounding of the lips necessary to produce the Swedish vowel /ɑ/ may result in a slight increase in perceived effort compared to /a/, as it requires activation of the orbicularis oris muscles around the mouth (Engstrand, 2004). This freedom to interpret the vowel sound could lead to unexpected differences not only between Swedish and French test results, but also within the Swedish study, possibly even within individual subjects. Had instructions on the pronunciation of the vowel been included in the pronunciation practice sheet, potential pitfalls such as these might have been reduced.

Implications for future studies

The contrasts explored in the present study are by no means exhaustive. An example of a parameter that would be interesting to investigate in Swedish is the aforementioned aspiration with regard to voiceless plosives. It would be worthwhile to explore this particular contrast to see whether the presence or absence of aspiration affects perceived vocal and articulatory effort, particularly in light of Robieux's hypothesis regarding Swedish. However, the test material used in the present study was developed for native French speakers and thus not designed with this specific contrast in mind. Prior to the start of the experiment, speech sounds present in French and not in Swedish had been excluded from the test forms developed by Robieux. However, no items were added to accommodate for Swedish speech sounds. Furthermore, analysis of data for the voiced-voiceless contrast made no distinction between plosives and fricatives. It is therefore not appropriate to draw conclusions about aspiration within the limits of the present study. The contrast aspirated-unaspirated voiceless plosives would be a valuable addition to a future test on Swedish subjects.

The present study is still at an early stage of development, yet opens up possibilities for additional research on vocal effort, and how voice therapy could be adapted according to subjective experiences. It is worth stating that a high level of perceived vocal and articulatory effort is not necessarily something to avoid during voice therapy. Rather, the findings offer some insight into how to structure therapy according to a patient's personal needs.

The general format lends itself to replication for other languages. The test could be adapted for native speakers of other languages and results of these languages correlated. Documented similarities between results for different languages could further encourage voice therapists to share ideas internationally in order to increase their knowledge and range of therapeutic exercises.

A range of information was gathered on participants, including gender, education level, presence of asthma, history of smoking, and further languages spoken. Another parameter that was not taken into account was participants' awareness of vocal function and their own voice. It would be interesting and worth exploring possible correlations between some of these variables and self-perception of vocal effort.

References

Carlsson, I, Eriksson, C, Persson, E L, Svanberg, P (1985). *Rösthantverket - en idébank för kliniskt erfarna röstlogopeder*. (Dissertation) Karolinska Institute, institution of phoniatrics and speech pathology, Huddinge sjukhus, Sweden. (Swedish)

Engstrand, O (2004). *Fonetikens grunder*. Sweden: Studentlitteratur, 121. (Swedish)

Engstrand, O (1999). Illustrations of the IPA: Swedish, in *Handbook of the International Phonetic Association*. United Kingdom: Cambridge University Press, 140-142.

Fant, G (1960). *Acoustic theory of speech production.* Netherlands: Mouton Publishers, 15-17.

Gay, T (1977). Articulatory movements in VCV sequences. *Journal of the Acoustic Society of America*, 62, 183-193.

Grillo, E U, Perta, K, & Smith, L (2009). Laryngeal resistance distinguished pressed, normal, and breathy voice in vocally untrained females. *Logopedics Phonatrics Vocology*, *34*(1), 43-48.

Johnson, K (2011). *Acoustic and auditory phonetics*. Chichester: Wiley-Blackwell, 25, 156.

Ohala, J (1983). The origin of sound patterns in vocal tract constraints. *The production of speech,* 189-216.

Rosenthal, A L, Lowell, S Y, & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice*, *28*(2), 144-153.

Robieux, C (2015). Effect of voicing on the self-perception of effort in French consonants production. (Accepted for publication in *Proceedings of the ICPhS 2015*, Glasgow, UK).

Van den Berg, J (1958). Myoelastic-aerodynamic theory of voice production, *Journal of Speech and Hearing Research* 3(1): 227-244.

ASHA (American speech-language-hearing association) (2005). Compendium of EBP Guide-line and Systematic Reviews. Retrieved 25/3-15. www.asha.org/members/ebp/compendium/

Temporal aspects of breathing and turn-taking in Swedish multiparty conversations

Jonna Hammarsten[1], Roxanne Harris[1], Nilla Henriksson[1], Isabelle Pano[1], Mattias Heldner[2] and Marcin Włodarczak[2]
[1] CLINTEC, Division of Speech and Language Pathology, Karolinska Institutet, Stockholm
[2] Department of Linguistics, Stockholm University, Stockholm

Abstract

Interlocutors use various signals to make conversations flow smoothly. Recent research has shown that respiration is one of the signals used to indicate the intention to start speaking. In this study, we investigate whether inhalation duration and speech onset delay within one's own turn differ from when a new turn is initiated. Respiratory activity was recorded in two three-party conversations using Respiratory Inductance Plethysmography. Inhalations were categorised depending on whether they coincided with within-speaker silences or with between-speaker silences. Results showed that within-turn inhalation durations were shorter than inhalations preceding new turns. Similarly, speech onset delays were shorter within turns than before new turns. Both these results suggest that speakers 'speed up' preparation for speech inside turns, probably to indicate that they intend to continue.

Introduction

In a conversation people exchange roles from speaker to listener on a regular basis. In order for turn-taking to proceed smoothly people use various cues – both conscious and unconscious ones (Rochet-Capellan et al., 2014). By means of different verbal and nonverbal cues, conversation partners can show the intention to take, hold or release the turn. These cues include, among other things, syntax, prosody and communicative silences (Local & Kelly, 1986) as well as head and body movements (Hadar et al., 1983, 1984).

From a physiological point of view, the foundation of speech production lies within the respiratory patterns of inhalations and exhalations. A great majority of speech sounds are formed when air is forced from the lungs via the glottis, through the oral and nasal cavities, where they become audible (Ohala, 1990). A typical breathing pattern during speaking consists of short and fast inhalations, followed by extended exhalations (Hixon et al., 1973).

Even though breathing has been mentioned as a potentially important turn-taking cue, few empirical studies have explored this area. However, according to Ishii et al. (2014), speakers' inhalations can be effective signals for predicting whether the turn will be released or kept. Ishii et al. (2014) also found that listeners' inhalations can predict attempts to take the turn. Similarly, results in Rochet-Capellan et al. (2014) suggest that most successful turns are initiated right after a new inhalation, and that breathing cycles at the start of a turn are more symmetric than inside a turn. Furthermore, Rochet-Capellan et al. (2014) found that speech onset delay (the interval between exhalation onset and speech onset) was kept to a minimum in successfully taken turns.

In line with the findings of Ishii et al. (2014) and Rochet-Capellan et al. (2014), the current study explores whether speakers show the intention to hold the turn using inhalation duration and speech onset delay as cues. We hypothesise that both inhalation durations and speech onset delays will be shorter within turns than when new turns are initiated.

Method

Participants

Participants were recruited by email sent to students at Stockholm University and KTH, Royal Institute of Technology. In this study two conversations were chosen for analysis. In one of the conversations, the speakers were a brother, aged 24, a sister, aged 28 and their father, aged 67. The other conversation included three students who did not know each other; a

female aged 23, a female aged 24 and a male aged 27. All participants were native Swedish speakers. The participants were not fully aware of the purpose of the study or the details of interest.

Procedure

The participants were instructed to converse freely in Swedish for about 20 minutes. The recordings were made in a sound treated room in the Phonetics Laboratory at Stockholm University. They were standing in an upright position around a table of 95 cm height. This position was preferred to minimize disruption of the breathing signal. The sound was recorded with close-talking directional microphones, (Sennheiser HSP 4). Respiratory Inductance Plethysmography (Watson, 1980) was used to measure respiratory activity. Each participant wore two elastic belts – one around the ribcage at armpit level and one around the abdomen, at the navel (Figure 1). Coils embedded in each belt are sensitive to changes in cross-sectional area due to breathing. The belts were connected to a RespTrack, a respiratory belt processor designed and built at Stockholm University (Edlund et al., 2014).

The respiratory signal from each belt was weighted and summed by means of a potentiometer, also part of RespTrack. The resulting signal (Figure 2) was captured by an integrated physiological data acquisition system (PowerLab by ADInstruments).

Figure 1. Students in the Phonetics Laboratory wearing transducer belts, each connected to a RespTrack, during a recording session.

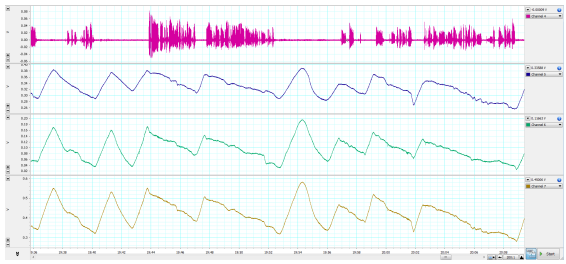

Figure 2. Respiratory signals shown as one wave for each transducer belt together with their summed wave below. The speech signal at the top.

The following parameters were estimated before the recording for each speaker: vital capacity (maximum exhalation produced after maximum inhalation), isovolume manoeuver (net lung volume change) (Konno & Mead, 1967), and resting expiratory level (measured after a relaxed sigh).

Annotation and feature calculation

Stretches of speech and silence were detected automatically in ELAN (Wittenburg et al., 2006) based on the intensity threshold for a silent portion of the signal, and then corrected manually. Next, inhalations and exhalations in the respiratory cycles were labelled automatically using an algorithm described in Włodarczak & Heldner (in press). This was also followed by manual correction in Praat (Boersma & Weenink, 2015).

Silent portions were classified as within-speaker silences (WSSs) or between-speaker silences depending on whether the same speaker or a new speaker continued the conversation after the silence (Jaffe & Feldstein, 1970).

Inhalation duration was calculated from the manually corrected respiratory annotations. Speech onset delay was calculated from the offset of the inhalation to the onset of speech. Cases were speech onset preceded inhalation offset were excluded. Overlaps as well as stretches of speech shorter than 1 second, which correspond primarily to short feedback expressions (Heldner et al., 2011) were not included in the analysis.

The final analysed sample contained 68 WSSs and 138 BSSs. Data analysis was done in SPSS, version 22.0 (IBM Corp, 2012).

Results

Comparison of mean parameter values in the two interval types shows that inhalation in

WSSs are shorter by 0.345 s (cf. Table 1), indicating that speakers inhale more quickly when they intend to hold the turn. A mean difference was also observed with respect to speech onset delay, whose duration was shorter by 0.356 s in WSSs than in BSSs (cf. Table 2). Thus, speakers tend to start speaking more quickly after the inhalation offset when holding the turn.

Histograms of inhalation duration and speech delay showed the distributions in the sample were positively skewed (cf. left panels in Figures 3 & 4). In order to remove the skew, the values were transformed using logarithms with a base of 10 (cf. right panels of Figures 3 & 4). The variations in log-transformed inhalation duration and speech onset delay of WSSs and BSSs are summarized in boxplots in Figure 5.

The log-transformed distributions satisfied the preconditions of an independent parametric t-test. With degrees of freedom corrected due to unequal variances between conditions (see Tables 1 and 2), significant differences between WSS and BSS categories were obtained both for inhalation duration ($t_{117.003} = 6.432$, $p < 0.0001$) and speech onset delay ($t_{202.883} = 3.428$, $p < 0.001$). The t-test thus revealed significant differences between WSSs and BSSs for both conditions in the current sample.

Table 1. Mean values and standard deviation of inhalation duration (in seconds) for Within Speaker Silences (WSSs) and Between Speaker Silences (BSSs).

Interval type	Mean	Std. Dev.
WSS	0.540	0.272
BSS	0.885	0.445

Table 2. Mean values and standard deviation of speech onset delay (in seconds) for Within Speaker Silences (WSSs) and Between Speaker Silences (BSSs).

Interval type	Mean	Std. Dev.
WSS	0.116	0.088
BSS	0.472	0.976

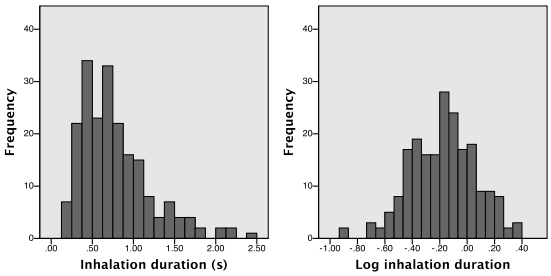

Figure 3. Histograms of inhalation durations (s) in raw values (left panel) and after log-transformation (right panel).

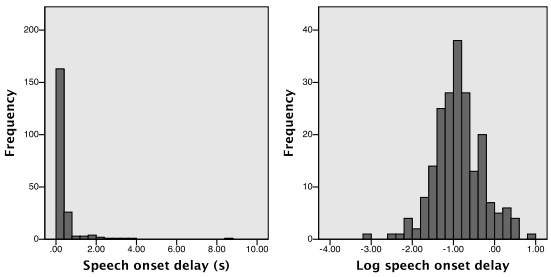

Figure 4. Histograms of speech onset delays (s) in raw values (left panel) and after log-transformation (right panel).

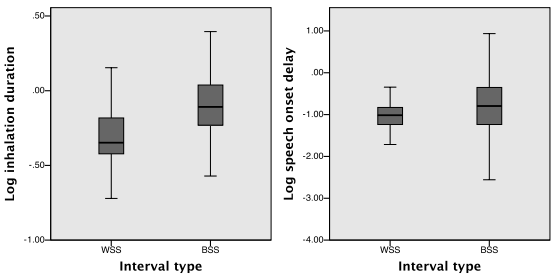

Figure 5. Boxplots of log-transformed inhalation duration (left panel) and speech onset delay (right panel) for the two interval types: Within Speaker Silences (WSSs) and Between Speaker Silence (BSSs).

Discussion

The aim of the study was to investigate whether inhalation duration and speech onset delay differs between turn-holding and turn-taking. In line with our hypothesis and earlier findings by Rochet-Capellan et al. (2014) and Ishii et al. (2014), the results showed significant differences between the conditions with lower values for both inhalation and speech onset delay within a turn compared to when a new turn was initiated.

The differences observed in inhalation duration were expected since the speaker holding the turn might want to keep his/her

inhalation to a minimum in order not to lose the turn. In a similar vein, Rochet-Capellan et al. (2014) suggested that speakers will adapt their breathing in order to hold the turn. Specifically, more rapid inhalations within a turn will reduce pauses and increase speaker's chances to continue without interruption.

Differences observed in speech onset delay were also expected, and for a similar reason. Specifically, the speaker might want to start speaking faster upon exhalation onset while holding the turn to reduce pause duration and counteract a possible interruption attempt from a dialogue partner.

Since the results are based on a small sample, containing in total 6 participants and 40 minutes worth of conversation, they need to be treated as preliminary. The environment of the phonetics laboratory in which the study took place could have affected conversational flow. Nevertheless, the material used in the present study was significantly more natural and spontaneous than the dialogues used by Rochet-Capellan et al. (2014) and Ishii et al. (2014). In spite of less tight experimental control similar results were obtained, indicating that the current setup allows for studying respiration in spontaneous multiparty interactions.

Despite the significance of the presented results, they can only be applied to the current sample. Future studies within the field of breathing and turn-taking are needed. They should include a larger amount of speakers as well as a larger number of conversations analysed, to ensure generalizability of the results and to the lay foundation for technological advances, such as dialogue systems and synthetic speech.

Acknowledgements

This work was funded in part by the Swedish Research Council project 2014-1072 *Andning i samtal* (*Breathing in conversation*).

References

Boersma P and Weenink D (2015). Praat: doing phonetics by computer [Computer program] (Version 5.3.84). Retrieved from http://www.praat.org/

Edlund J, Heldner M and Włodarczak M (2014). Catching wind of multiparty conversation. In: J Edlund, D Heylen & P Paggio, eds, *Proceedings of Multimodal Corpora: Combining applied and basic research targets (MMC 2014)*. Reykjavik, Iceland.

Hadar U, Steiner T, Grant E C and Rose F C (1983). Head movement correlates of juncture and stress at scentence level. *Language and Speech* 26, 117-129.

Hadar U, Steiner T, Grant E C and Rose F C (1984). The timing of shifts of head postures during conversation. *Human Movement Science* 3, 237-245.

Heldner M, Edlund J, Hjalmarsson A and Laskowski K (2011). Very short utterances and timing in turn-taking. In *Proceedings Interspeech 2011*. Florence, Italy, 2837-2840.

Hixon T J, Goldman M D and Mead J (1973). Kinematics of the chest wall during speech production: Volume displacement of the rib cage, abdomen, and lung. *Journal of Speech, Language and Hearing Research* 16, 78-115.

IBM Corp. (2012). SPSS Statistics for Macintosh [Computer program] (Version 22.0). Armonk, NY: IBM Corp.

Ishii R, Otsuka K, Kumano S and Yamato J (2014). Analysis of respiration for prediction of "who will be next speaker and when?" in multi-party meetings. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, 18-25.

Jaffe J and Feldstein S (1970). *Rhythms of dialogue*. New York, NY, USA: Academic Press.

Konno K and Mead J (1967). Measurement of the separate volume changes in the rib cage and abdomen during breathing. *Journal of Applied Physiology* 22, 407-422.

Local J K and Kelly J (1986). Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies* 9, 185-204.

Ohala J J (1990). Respiratory activity in speech. In: W J Hardcastle & A Marchal, eds, *Speech Production and Speech Modelling*: Springer, 23-53.

Rochet-Capellan A, Bailly G and Fuchs S (2014). Is breathing sensitive to the communication partner? In: N Campbell, D Gibbon & D Hirst, eds, *Proceedings of Speech Prosody 2015*. Dublin, Ireland: Trinity College, 613-617.

Watson H (1980). The technology of respiratory inductive plethysmography. In: F D Stott, E B Raftery & L Goulding, eds, *Proceeding of the Second International Syposium on Ambulatory Monitoring (ISAM 1979)*. London: Academic Press.

Wittenburg P, Brugman H, Russel A, Klassmann A and Sloetjes H (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* 1556-1559.

Włodarczak M and Heldner M (in press). Respiratory properties of backchannels in spontaneous multiparty conversation. In *Proceedings ICPhS 2015*. Glasgow, UK.

Deaccented verbs in Swedish

Anna Hed and Anna Smålander
Department of Linguistics and Phonetics, Lund University

Abstract

Deaccenting of verbs in Swedish has previously been briefly described in the literature. It has been said to occur in verbs in particle verb constructions, lexicalized phrases, before indefinite plural objects and some specific constructions. This study shows a wider variety of deaccented verbs. It shows that it can take place in the whole particle verb construction before another constituent, before definite singular objects as well as before prepositional phrases functioning as adverbials.

Introduction

In this article, we present a small study on deaccenting of verbs in Swedish. We look at some instances where the verb is deaccented before and incorporated into a prosodic word together with its object or following constituent, much like a particle verb, where the verb is deaccented before its particle. The phenomenon is briefly described in previous research.

Myrberg & Riad (2013) mention deaccenting of content words. According to them, deaccenting occurs in three types of phrases (p. 260):

1. lexicalized phrases such as *röda MATtan* 'red carpet'
2. some names such as *Röda KORSet* 'Red Cross'
3. particle verbs such as *hoppa UPP* 'jump up'

In all these cases, the non-final word is deaccented. Myrberg & Riad (2013) state that these phrases look much like compounds, since they behave as what they call one maximal prosodic word. In their account, the maximal prosodic word can contain only one accent. The difference in relation to compounds is that the stress is to the right in deaccented phrases, whereas it is to the left in compounds. Since the pitch accent associates with the head, which in these cases is the second word, the initial word is deaccented in the three cases mentioned. According to Riad (2012), the reason why it is possible to have deaccented lexical words in running speech is because culminativity is non-obligatory at the level of the maximal prosodic word in Swedish.

One phenomenon not mentioned by Myrberg & Riad (2013) is deaccenting of verbs that are not part of particle verb constructions. This kind of deaccenting is mentioned by Myrberg in her dissertation about the intonational phonology of Stockholm Swedish (2010). She describes deaccenting in particle verbs and lexicalized phrases, such as *hänga LÄPP* 'be sad', lit. 'hang lip' and *spela KORT* 'play cards' (p. 81). However, she also states that verbs (including auxiliary verbs) can be deaccented when they are followed by a plural indefinite object (p. 82), as in sentence 4.

4. *Jag ska laga kopiaTORer*.
   'I'm going to mend copying machines'
   (Myrberg, 2010, p. 82)

Here the verbs *ska laga* are deaccented. She also points out that this applies to whole particle verb constructions as well, as in 5.

5. *Hon bryter sönder SAXar*.
   'She breaks scissors'
   (ibid.)

Here *bryter sönder* is deaccented. Myrberg notes that the scope and constraints of this phenomenon have not been thoroughly studied.

Anward & Linell (1975) studied lexicalized phrases, which they found to receive a 'summarizing accent' (*sammanfattningsaccent*), where the non-final word is deaccented. However, they also mention syntactic constructions where the object is replaceable by another object with a similar semantic meaning, which can receive "sammanfattningsaccent". Some examples are 6 and 7.

6. *dricka VIN (öl, kaffe, vatten)*
   'drink wine (beer, coffee, water)'
   (Anward & Linell, 1975, p. 107)
7. *lukta BRÄNT (skit, mat, gott)*
   'smell [like] burnt (shit, food, good)
   (ibid.)

Here the non-final word is deaccented.

The aim of this study is to describe different types of deaccenting of verbs in Swedish, both those types which have been described in previous research, which can be corroborated by our empirical data, and other types that are present in our material, which to the best of our knowledge have not been touched upon yet. We do not present a thorough analysis of the phenomenon, but point to some interesting findings worthy of further investigation.

Material and method

The material for this study was not recorded for the specific purposes of the present research question, but to be used in a game application. This means that the material is not controlled to show the full range of how deaccenting of verbs works, but it will provide a hint for further investigation. The persons recorded were between 19-32 years old and they were all from the Stockholm Region. There were 4 men and 2 women and in total 800 different sentences was recorded. 600 sentences differentiated between non-focal nouns with accent 1-associated noun suffixes (the definite singular suffix *-en/-et*) and with accent 2-associated noun suffixes (the indefinite plural suffix *-er/-ar*), as in example 8. 200 sentences differentiated between non-focal verbs with accent 1-associated verb suffixes (the present tense suffix *-er*) and with accent 2-associated verb suffixes (the past tense suffix *-te*), as in example 9.

8. a. *Anna såg falken i trädet.*
'Anna saw the falcon in the tree.'
b. *Anna såg falkar i trädet.*
'Anna saw falcons in the tree.'

9. a. *Solen bleker Lovisas kläder.*
'The sun bleaches Lovisa's clothing.'
b. *Solen blekte Lovisas kläder.*
'The sun bleached Lovisa's clothing.'

When going through the material, we discovered several cases where the verbs were deaccented. We decided to look further into this phenomenon. We went through all the material, and noted every instance of deaccenting. The sentences were then classified into different syntactic categories. F0 curves were extracted in Praat.

Results

In our material, we found instances of deaccenting in both lexicalized phrases and in particle verbs, as would be expected according to Myrberg & Riad's categorization. The final constituents in these cases were associated with a word accent, as is expected. One example of a deaccented verb in a lexicalized phrase occurred in the following sentence:

10. *Gabriel stöper ljus på julen.*
'Gabriel makes candles at Christmas.'

Stöper is deaccented and *ljus* carries the word accent. The deaccenting also occurred in the corresponding past tense sentence.

An example of deaccenting of a verb in a particle verb construction is shown in sentence 11.

11. *Charlotta fryser in mycket bär på sommaren.*
'Charlotta freezes a lot of berries in the summer.'

Fryser is deaccented and *in* carries the word accent.

We also found the kind of deaccenting that Myrberg mentions in her dissertation, the kind where a verb is deaccented when followed by an indefinite plural noun phrase. An example is shown in sentence 12.

12. *Amanda planerade rutter med hjälp av GPS.*
'Amanda planned routes with GPS.'

Here, the speaker deaccented the verb *planerade* only in this case, not in the corresponding definite singular noun phrase sentence, *Amanda planerade rutten* ('the route') *med GPS*. This is shown in figure 1 and 2.

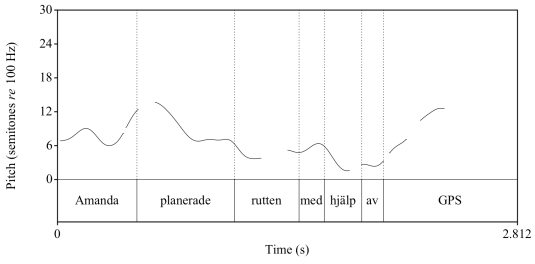

Figure 1. 'Amanda planned the route with GPS.'

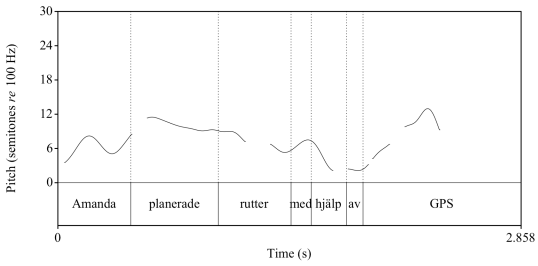

Figure 2. 'Amanda planned routes with GPS.'

There were also deaccented particle verbs, which is also briefly touched upon by Myrberg (2010, p. 82). This is shown in figure 3.

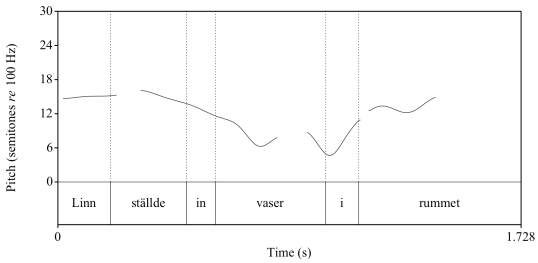

Figure 3. 'Linn put vases in the room.'

Here, the whole particle verb construction ställde in is deaccented, not just the verb ställde.

Furthermore, in VPs with auxilliary verbs, both the auxiliary and the main verb can be deaccented, as is shown in figure 4, where *måste byta* is deaccented. This example can be compared to sentence 4 above.

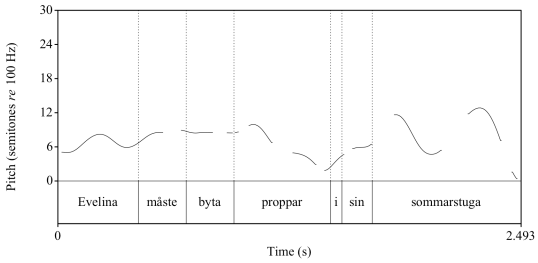

Figure 4. 'Evelina has to change fuses in her summer house.'

Apart from these more expected findings, which corroborate previous research on the phenomenon, we found three types of deaccenting that to the best of our knowledge have not been mentioned in previous literature. First, deaccenting can occur when the object is not indefinite plural, but definite singular, as in figure 5.

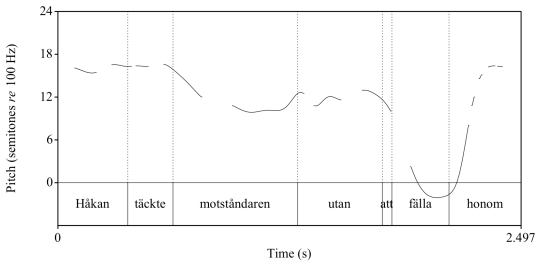

Figure 5. 'Håkan covered the opponent without tripping him.'

Here, the phrase is not a lexicalized phrase, so the deaccenting of the verb *täckte* cannot be caused by that. The verb is not a particle verb, and, as mentioned, the noun phrase is definite singular, not indefinite plural.

Second, not only objects can follow the deaccented verbs but also PPs, where the PP functions as an adverbial, as in figure 6.

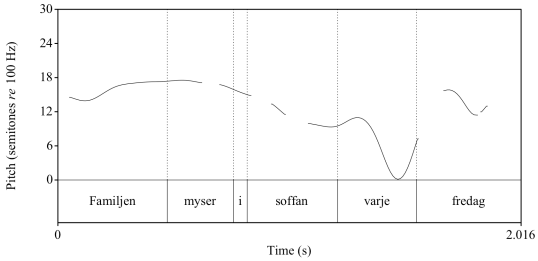

Figure 6. 'The family cuddles in the sofa every Friday.'

The third type is a mixture of some previously mentioned types. They bear similarities to the type that Anward & Linell (1975) describe, where some verbs are prone to be deaccented in similar constructions but with different objects, see sentence 6 and 7 above. Examples from our material, with the verb *satte* 'put', are shown in figure 7, 8 and 9.

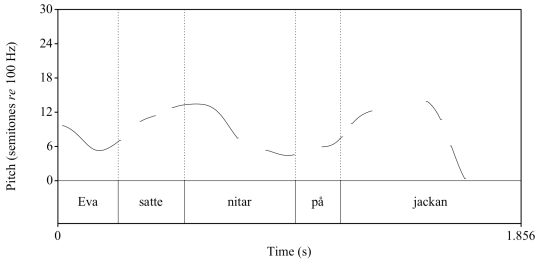

Figure 7. 'Eva put rivets on the jacket.'

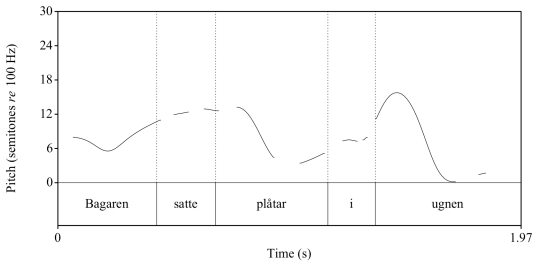

Figure 8. 'The baker put trays in the oven.'

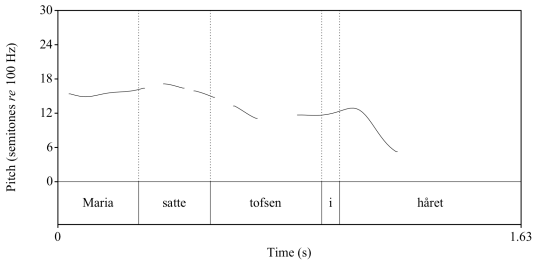

Figure 9. 'Maria put the hair-tie in the hair.'

In the first two examples above, the objects are indefinite plural, and notably, in the third one, the object is definite singular. The verb *satte* seems to be easily deaccented. The difference between these examples and those in Anward & Linell is that their examples are only indefinite singular cases, whereas in our material, both indefinite plural and definite singular seem to be able to be accompanied by deaccenting.

When it comes to particle verbs, there is one more interesting finding. To make sure that our speakers would put accents on the verbs we needed to record, adverbials were inserted between the verb and the particle, since we assumed that the speakers then might treat the constituents as separate prosodic words. However, the speakers had a few different strategies to mark particle verbs nevertheless. Figure 10 shows an example where the verb is deaccented and forms a prosodic word together with the adverbial, which receives the word accent.

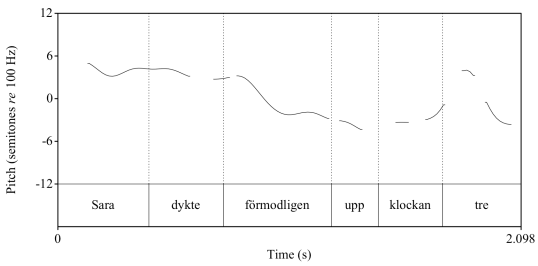

Figure 10. 'Sara probably showed up at three.'

This type is mentioned by Myrberg & Riad (2013) as well.

Another strategy is shown in figure 11, where the speaker forms one prosodic word of the verb, the adverb and the particle. The particle carries the word accent.

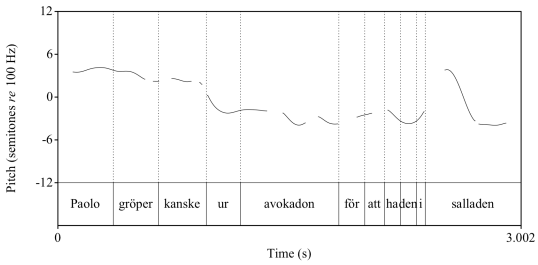

Figure 11. 'Paolo maybe scoops out the avocado to put it in the salad.'

Finally, figure 12 shows how the verb forms a prosodic word on its own, but the adverbial *kanske* 'maybe' gets deaccented before the particle.

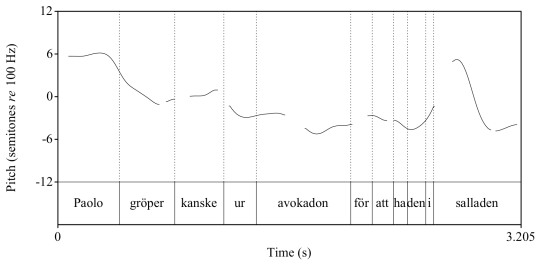

Figure 12. 'Paolo maybe scoops out the avocado to have it in the salad.'

Discussion

As can be seen in the data, the whole VP, or the VP and a PP, forms one prosodic word with only one word accent, which in these cases always is located on the NP/PP (the object/adverbial). This deaccenting has been previously described in relation to Swedish word accents, but our material showed a wider range of occurrence of the phenomenon than has hitherto been described.

As stated, our purpose is not to provide an exhaustive analysis, but to point at new research possibilities. This phenomenon needs to be studied more closely and more systematically in order to get a full picture of how deaccenting works. We do not know how common the phenomenon is, since we cannot provide statistical data from this study, as the material was recorded for another purpose. Another question concerns what the motivation for deaccenting is. Is there any pragmatic or semantic reason to deaccent the verbs? Are there verbs that can never be deaccented and in that case, why?

Summary

This article has shown a wide range of cases where verbs are deaccented. We have both corroborated previous research findings by e.g. Myberg & Riad and Anward & Linell with empirical data, and presented new cases of deaccenting of verbs. A systematic analysis of these other types of deaccenting is desirable and remains to be undertaken.

Acknowledgements

This work was supported by Marcus and Amalia Wallenberg Foundation (grant number 2014.0039).

References

Anward J, Linell P (1975). Om lexikaliserade fraser i svenskan. *Nysvenska studier*, 77-119.

Myrberg S (2010). *The intonational phonology of Stockholm Swedish*. Stockholm studies in Scandinavian philology, 53. Stockholm University.

Myrberg S, Riad T (2013). The prosodic word in Swedish. In: Asu E L, Lippus P, eds, *Proceedings of Nordic Prosody XI*, 255-264.

Riad T (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, 122: 1352-1379.

A change in the openness of two vowel phoneme pairs in eastern Icelandic: An acoustic analysis of /œ:/ vs. /ʏ:/, and /ɛ:/ vs. /ɪ:/

Guðlaug Hilmarsdóttir
Centre for Languages and Literature, Lund University

Abstract

The number of Icelanders lacking the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/ is decreasing. Since Guðfinnsson's study (1964), and up to Árnason and Þráinsson's study (Árnason and Pind 2005), speakers found in the capital region, and in eastern Iceland, are maintaining the distinction between the rounded, and between the unrounded vowel phonemes to a greater degree. This study serves to test the proposition from the previous researches. Overall results show that the participants in eastern Iceland have a mixed pronunciation, i.e. the participants partly lack the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/, and partly maintain it. The mergers result in sounds close to /e:/ and /ø:/ respectively. This was also the case with the speakers from the capital region. Overall, the lack of distinction was less apparent amongst the young speakers from the East than in the capital region.

Introduction

Icelandic has several dialect features, found in different regions of the country. These are features such as vowel phoneme mergers, diphthongization of monophthongs, monophthongization of diphthongs, and harder, or softer, pronunciation of certain consonants.

The number of Icelanders lacking the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/, however, is decreasing. Since Guðfinnsson's (1964) study, and up to Árnason and Þráinsson's (Árnason and Pind, 2005) study, speakers found in the capital region, and in eastern Iceland, are moving towards a general pronunciation, i.e. more speakers are maintaining the distinction between the two unrounded vowel phonemes, and between the two rounded vowel phonemes.

This study serves to test the proposition from the previous researches, with the following research question:

> Do young speakers in eastern Iceland lack the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/, as has been found in the capital region in previous research?

This is an empirical acoustic investigation with a comparative approach, which contributes phonological data to the field of general linguistics and Icelandic dialectology. The quality of the two vowel phonemes within each pair are compared between eastern Icelandic and the Icelandic found in the capital region. It is expected that the young speakers in the East, and in the capital region, will lack the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/ in part. Also, it is expected that the young speakers in the capital region will lack this distinction to a larger degree than the young speakers in the East.

Background

Icelandic vowel phonemes and their acoustic properties

Vowel quality is described with whether they are front, back or central; and with the degree of opening, i.e. close, close-mid, open-mid, and open. Icelandic has eight vowel phoneme pairs as monophthongs:

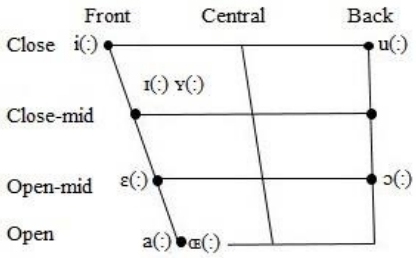

Figure 1. The eight Icelandic vowel phoneme pairs.

Figure 1 shows that the Icelandic vowel phoneme pairs spread on six locations in the vowel square. These locations are descriptive of the highest point of the tongue in the oral cavity. Figure 1 shows that Icelandic has six front vowel phoneme pairs, these being /i/ and /i:/, /ɪ/ and /ɪ:/, /ʏ/ and /ʏ:/, /ɛ/ and /ɛ:/, /a/ and /a:/, and /œ/ and /œ:/. Examples of these vowel phoneme pairs are in words such as *nýttir* (Eng. *made use of*) and *nýtir* (Eng. *make use of*), *fylla* (Eng. *fill*) and *fila* (Eng. *blanket*), *munnur* (Eng. *mouth*) and *munur* (Eng. *difference*), *vellur* (Eng. *boils*) and *velur* (Eng. *chooses*), *fatta* (Eng. *figure out*) and *fata* (Eng. *bucket*), and *völlur* (Eng. *field*) and *völur* (Eng. *pebbles*). Out of these, two vowel phoneme pairs are not as front, these being /ɪ/ and /ɪ:/, and /ʏ/ and /ʏ:/. Figure 1 also shows that Icelandic also has two back vowel phoneme pairs, where one is close, and the other open-mid, i.e. /ɔ/ and /ɔ:/, and /u/ and /u:/ respectively. Examples of the back allophone pairs are in words such as *hoppa* (Eng. *jump*) and *hopa* (Eng. *regress*), and *húkka* (Eng. *catch (a ride)*) and *húka* (Eng. *squat*).

Vowels are also described with the degree of openness of the oral cavity. Figure 1 shows that there are two close allophone pairs, two open mid, and two open. Then, two allophone pairs are halfway between being close and close-mid, i.e. /ɪ(:)/ and /ʏ(:)/. This also describes how high or low the highest point of the tongue is:

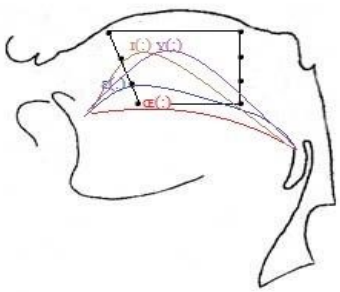

Figure 2. The degree of opening for /œ(:)/, /ɛ(:)/, /ɪ(:)/, and /ʏ(:)/.

Figure 2 shows how tongue position varies between the rounded and the unrounded vowel phonemes. Lindblad (2010: 93) notes that in general, a change in the form of the vocal tract can affect formant frequencies. However, different changes in the form of the vocal tract affects formant frequencies in a different manner. As an example, a decrease in the degree of openness, i.e. when the tongue moves higher up, can lead to a higher correlation between the first formant and the pharynx, and a higher correlation between the second and the third formant and the oral cavity. Nevertheless, Lindblad (2010: 93) notes that this cannot be generalized, since many vowels differ in their degree of openness.

However, Lindblad (2002: 94) offers five points as guidelines for analysing the relationship between formant frequencies and the form of the vocal tract:

1. Lip rounding or other hinders lower formant frequencies.
2. The smaller degree of openness, the lower the frequency of the first formant.
3. The closer the narrowest passage between the tongue and the roof of the oral cavity is to the mouth opening, the higher the frequency of the second formant, and also the third formant, but to a lesser extent.
4. The larger the degree of openness under the tongue, the lower the frequency of the third formant.
5. The longer the vocal tract, the lower the frequencies of the different formants.

These five points will guide the analysis of the formant frequencies in the results section.

Lindblad's (2002: 94) five points, together with Figure 2, can be used to predict what a merger between /œ:/ and /ʏ:/, and between /ɛ:/ and /ɪ:/, would result in, in terms of the change in the form of the vocal tract and the affect these changes have on formant frequencies.

In a merger within each vowel phoneme pair, a change to a less open vocal tract (/œ:/ to /ʏ:/, /ɛ:/ to /ɪ:/) would result in lower frequency values of the first formant (Lindblad's second point). On the contrary, a change to a more open vocal tract (/ʏ:/ to /œ:/, /ɪ:/ to /ɛ:/) would result in higher frequency values of the first formant (opposite of Lindblad's second point). The position of the narrowest passage of the tongue is responsible for the second formant value. As

both vowel phoneme pairs are considered front vowel phonemes, the effect of F2 variation will be included impartially.

Vowel phoneme mergers

One of the dialect features found in certain areas of Iceland, is the merger between /œ:/ and /ʏ:/, and between /ɛ:/ and /ɪ:/. Mergers occur over time, when two or more sounds that once were distinguished by speakers, merge, in a certain dialect or language, and therefore become one sound (Hickey, 2004: 125). Hickey further notes that the merged sounds can later move together to a different location on the vowel square, which is termed as a shift (2004: 125).

Vowel phonemes can be affected by other sounds in the context, both by consonants and other vowels phonemes (Árnason and Pind, 2005: 253). Hickey notes that vowel phoneme mergers, where the vowel phonemes are sensitive to context, are often determined by a following sonorant, i.e. /n, l, r/ (2004: 127). Furthermore, mergers of this type do not seem to be determined by a following obstruent, e.g. /t, k, p/. Hickey claims that the reason for the effect from sonorants is because how alike the quality of a sonorant is with that of a vowel (2004: 127). Also, "it is […] known that the coda sonorants tend to become absorbed into the nucleus of the syllable they occupy" (Hickey, 2004: 127).

The *flámæli* (Eng. *flayspeech*) dialect feature in Iceland is an example of spontaneous changes. Árnason and Pind note that flayspeech is a dialect feature where the close-mid vowel phonemes /ɪ/ and /ʏ/ become more open, or diphthongize to /ɪɛ/ and /ʏœ/, or /ɛ/ and /œ/ (2005: 254).

Methodology

The reading method was chosen over elicitation, even though the latter might be more descriptive of natural speech. By choosing the reading method, it was possible to control for the appearance of both vowel phonemes.

Structuring the sentences

The four phonemes, i.e. /ɛ:/, /œ:/, /ɪ:/, and /ʏ:/, served as the main guidelines in finding key words to build up sentences that the participants to read. The dictionary *Íslensk Orðabók* (1997) was used to find words with these vowel phonemes in stressed position. The words within each phoneme pair, i.e. within /ɛ:/ and /ɪ:/, and within / œ:/ and /ʏ:/, were almost identical, except for that single vowel phoneme. Example words are those such as, *sögu* (Eng. *story*), and *sugu* (Eng. *sucked*). Random sentences were created manually, written on separate pieces of paper, and laminated in order to prevent unnecessary noise in the recordings. The following is the final set of keywords that were used in constructing the sentence pairs:

Table 1. List of key words with the unrounded vowel phonemes in stressed position.

/ɛ:/	/ɪ:/
beðin	biðin
betur	bitur
pela	pila
dekillinn	dikillinn
fela	fila
fetað	fitað

Table 2. List of key words with the rounded vowel phonemes in stressed position.

/œ:/	/ʏ:/
sögu	sugu
röðullinn	ruðullinn
rösull	rusull
nötur	nutur
mösull	musull
föður	fuður
börur	burur

Participants

Twenty speakers, aged between 16-20 years old, participated in this research. Half of the participants came from eastern Iceland, which served as the focus group, while the other half came from the capital area, which served as a control group. Each group was balanced for gender.

Recording sessions

A microphone was connected to a laptop and the speakers were recorded with Praat. The participants came in to a room, one by one, and sat in front of the microphone with their arms rested in their lap under the table.

One of the sentences was put on the table, which the participants read, and then it was taken away and a new sentence was read. The participants read all sentences in one recording.

After the recording had finished, the participants filled in a set of background questions.

Analysis

The first three formants were measured manually in Praat, despite the fact that the third formant is irrelevant, in case something interesting would show up. However, the final measurements showed that the third formant was in fact irrelevant, as it was in most cases found at very similar frequencies between each vowel phoneme in each pair.

R was used to create scatterplots from the documented formant frequencies. In addition, R was used to calculate mean the mean frequency of each vowel phoneme within each group of speakers, in order see if the formants were found in close frequency range within each pair. The standard deviation was also calculated in R, in order to see how the formants in each vowel phoneme would scatter. If standard deviation will be high, then the formants will scatter over a larger area on the scatter plot.

Results

Figures 4(a)-4(d) show the results of the measurements, in Hz, when young male and female speakers from the capital region of Iceland, and from the East, read sentences containing key words with either the /œ:/ or the /ʏ:/ vowel phoneme. Furthermore, Tables 3(a)-3(d) show the mean frequency, and standard deviation, also measured in Hz.

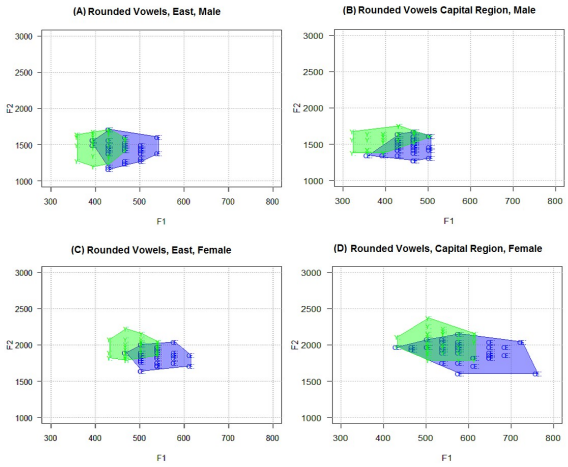

Figure 4. A scatter plot of the first two formants in /œ:/ and /ʏ:/.

Table 3. The mean frequency (MF) and the standard deviation (SD) of the first two formants in /œ:/ and /ʏ:/.

	F1(MF)	F1(SD)	F2(MF)	F1(SD)
(a) Male, East				
/œ:/	465.412	39.944	1439.000	132.717
/ʏ:/	409.618	32.258	1531.029	143.115
(b) Male, capital region				
/œ:/	457.606	31.243	1488.667	108.616
/ʏ:/	415.546	50.056	1576.970	101.974
(c) Female, East				
/œ:/	539.800	34.996	1850.857	91.146
/ʏ:/	494.857	35.191	1957.800	99.385
(d) Female, capital region				
/œ:/	584.886	31.243	1915.943	126.020
/ʏ:/	535.514	47.784	2029.686	139.659

Figures 4(a)-4(b), and Tables 3(a)-3(b) show that an overlap is found when the young male speakers from both regions pronounce both of the two rounded vowel phonemes. Tables 3(a)-3(b) show that the formants are found around similar frequencies for the young male speakers in the East and in the capital region. Standard deviation is relatively low in the case of both formants, which means the formants scatter over a smaller area on the scatter plot. Also, standard deviation of F2 in both rounded vowel phonemes is lower for the young male speakers in the capital region, which means that the second formant is found at a narrower frequency scale, i.e. the scatter of both rounded vowel phonemes is narrower on y-axis, than that of the young male speakers of the East.

Figures 4(c)-4(d), and Tables 3(c)-3(d) show that there is also an overlap when the young female speakers from both regions pronounce the two rounded vowel phonemes. Tables 3(c)-3(d) show that the first two formants in the rounded vowel phonemes are found at similarly close frequencies. Nevertheless, standard deviation of the first two formants in both rounded vowel phonemes is higher in the case of the young female speakers from the capital region, which means that the formants scatter over a greater area, at least in the case of F1. This allows for a greater overlap, as can be seen in Figure 4(d), compared with Figure 4(c).

Figures 5(a)-5(d) show the results of the measurements, in Hz, when young male and female speakers from the capital region of Iceland, and from the East, read sentences containing key words with either the /ɛ:/ or the /ɪ:/ vowel phoneme. Furthermore, Tables 4(a)-

4(d) show the mean frequency, and standard deviation, also measured in Hz:

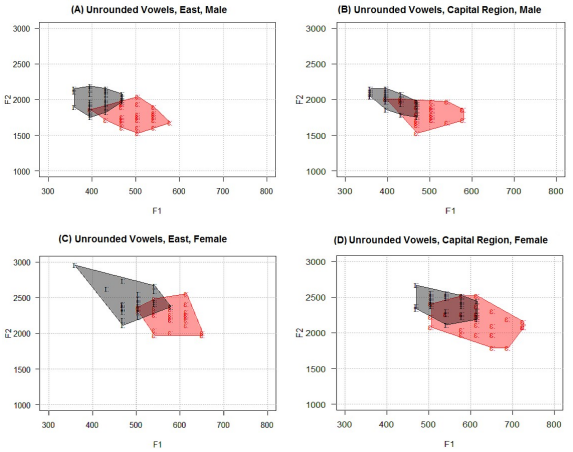

Figure 5. A scatter plot of the first two formants in all occurrences of the /ɛ:/ and /ɪ:/ vowel phonemes.

Table 4. The mean frequency (MF) and the standard deviation (SD) of the first two formants in /ɛ:/ and /ɪ:/.

Group	F1(MF)	F1(SD)	F2(MF)	F1(SD)
(a) ME				
/ɛ:/	495.857	39.638	1825.414	128.847
/ɪ:/	415.036	32.191	1998.207	122.661
(b) MC				
/ɛ:/	493.759	42.935	1488.667	122.871
/ɪ:/	413.207	38.598	1987.429	98.295
(c) FE				
/ɛ:/	579.621	41.546	1850.857	148.267
/ɪ:/	494.931	41.596	2436.138	167.253
(d) FC				
/ɛ:/	624.700	66.404	2162.033	182.437
/ɪ:/	539.267	45.546	2393.367	122.523

Figures 5(a)-5(d) show that the scatter fields for the /ɪ:/ vowel phoneme is of similar shape for all groups, except for the young male speakers in the East. However, despite that the scatter field for /ɪ:/ for the young female speakers in the East was of similar shape as for both groups in the capital region, the degree of scatter is nevertheless larger in that area. The scatter field for the /ɛ:/ vowel phoneme is of similar shape for the young male speakers from both regions. The scatter field for the /ɛ:/ vowel phoneme is also similar between the young female speakers in both regions. However, the degree of scatter is greater amongst the young female speakers in the capital region. This can be seen in how the field in Figure 5(d) extends over a longer frequency range, in terms of the first formant.

Figures 5(a)-5(b), and Tables 4(a)-4(b), show that an overlap was found when the young male speakers in the East, and in the capital region, were pronouncing the two unrounded vowel phonemes. Also, the first two formants are at equally close frequencies in both regions. Nevertheless, standard deviation of F1 is greater for young female speakers in the capital region, which means that the scatter of F1 in both unrounded vowel phonemes is greater than that of the young male speakers in the capital region. However, Tables 4(a)-4(b) show that standard deviation of F2 is greater for the young male speakers in the East, in both unrounded vowel phonemes

Figures 5(c)-5(d), and Tables 4(c)-4(d) show that an overlap was found when the young female speakers in the East, and in the capital region, were pronouncing the two unrounded vowel phonemes. Also, the first two formants in the unrounded vowel phonemes are found at similarly close frequencies in both regions. However, the second formant in the unrounded vowel phonemes is slightly further apart in the case of the young female speakers from the capital region. In addition, standard deviation is greater in almost all cases for the young female speakers in the capital region, except for F2 in the East. This allows for a greater overlap between the two unrounded vowel phonemes amongst the young female speakers from the capital region.

Discussion

Several interesting points should be discussed at this point. The lack of distinction between the vowel phonemes in both pairs is slightly greater amongst the young male speakers than for the young female speakers in the capital region. This might indicate different attitudes towards the overlap of the unrounded vowel phonemes, and the unrounded vowel phonemes.

In addition, the young female speakers in both regions have a greater scatter in the case of the unrounded vowel phonemes than that of the young male speakers, while in the case of the rounded vowel phonemes, the young male speakers in both regions have a greater scatter, except in the case of the first formant in /ʏ:/. This indicates that gender might matter in the usage of the two mergers

The fact that the overlap is only apparent in some example sentences questions the fact that the flayspeech dialect feature was is in fact an

example of spontaneous changes. Various consonants appeared in the following context of the rounded and the unrounded vowel phonemes, such as plosives, fricatives, and liquids. These different types of consonants might affect the young speakers in making a distinction between two vowel phonemes in different ways.

Neither of the more open vowel phonemes had completely moved to the more close vowel phonemes, nor did the opposite movement occur for the more close vowel phonemes. Thus the vowel phonemes met in the middle, forming sounds similar to /ø:/ and /e:/ respectively. This also means that for those young speakers who lack the distinction, the form of the vocal tract is less varied compared to those young speakers who do make the distinction. For both the rounded and the unrounded pair, the more open and the more close vowel phonemes have become a close mid vowel phoneme. Assumptions about tongue position are made based on formant measurements, as shown in the tables in the results, in reference to Lindblad's second and third point (2002, see Section 2.1).

As in Guðmundsson (1964), and in the RÍN research (Árnason and Pind, 2005), flayspeech was found in the same three counties in this study. This suggests that the results in Árnason and Þráinsson's study might not be entirely true (Árnason and Pind, 2005).

As Section 2.4.1 shows, the use of the flayspeech dialect feature had drastically decreased from Guðfinnsson's research in 1941-1943 until the RÍN research in the 1980s. The young speakers in the capital region had developed a new version (Árnason and Pind, 2005: 402), and the lack of distinction found amongst the speakers in this group is partly maintained in that region, and now found in the East of Iceland.

Conclusion

The analysis of the results have now shown that the research question stated in Section 1 has been answered. Previous research has shown that Icelanders are moving towards a more general pronunciation, where the distinction between /ɛ:/ and /ɪ:/, and between /œ:/ and /ʏ:/ is maintained. The current study shows otherwise.

The fact that this merger was found in the two regions indicates that either its usage is decreasing at a slower rate than indicated by Árnason and Þráinsson's RÍN research, or the generation of speakers that the participants in this study belong to, have started to use it, on purpose or not, and might do so in the coming future. Further study in the future is needed to predict the future of this dialect feature, whether its usage is in fact coming to an end, or is regaining popularity and increasing in usage.

Despite the fact that the context the vowel phonemes appeared in was not taken into consideration in the analysis of the results of this study, it raises the question of the flayspeech dialect feature to be a free variation or context sensitive, as this dialect feature was only partly apparent. This is what will be taken up in a master's thesis, in addition to taking the differences found between individual results and general results into consideration.

Dialect features are often closely connected with social aspects, such as attitudes, and also with the context the sounds in question appear in. Flayspeech is no exception. To be precise, "flayspeech is a little more complicated than swapping out *i*/*e*, and *u*/*ö*" (Hermannsdóttir, 2015: personal communication). Even though this this study indicated that the flayspeech dialect feature is in usage at this point in time, it is not enough to fully understand its past, present, or its future. However, future studies suggested in this section will give a clearer comparison of the current situation within the age group that is tested, which will make generalizations and future predictions easier

References

Árnason K, Pind J (2005). Hljóð. In: *Íslensk Tunga: Handbók um hljóðfræði og hljóðkerfisfræði*. Reykjavík: Almenna Bókafélagið.

Böðvarsson Á (1997). *Íslensk Orðabók*. 2d ed. Reykjavík: Mál og Menning.

Guðbjörnsson B (1964). *Mállýzkur II: Um Íslenzkan framburð*. Heimspekideild Háskóla Íslands og Bókaútgáfa Menningarsjóðs.

Hickey R (2004). Mergers, near-mergers and phonological interpretation. In Christian J Kay, Carole Hough and Irené Wotherspoon (eds.). *New Perspectives on English Historical Linguistics*. Amsterdam: John Benjamins.

Lindblad P (2002). *Grundläggande Akustisk Fonetik.* Lund: Lunds Universitet.

Lindblad P (2010). *Fonetikens Grunder*. Lund: Lunds Universitet.

On the temporal domain of co-speech gestures: syllable, phrase or talk spurt?

David House, Simon Alexanderson and Jonas Beskow
Department of Speech, Music and Hearing, KTH, Stockholm, Sweden

Abstract

This study explores the use of automatic methods to detect and extract hand gesture movement co-occuring with speech. Two spontaneous dyadic dialogues were analyzed using 3D motion-capture techniques to track hand movement. Automatic speech/non-speech detection was performed on the dialogues resulting in a series of connected talk spurts for each speaker. Temporal synchrony of onset and offset of gesture and speech was studied between the automatic hand gesture tracking and talk spurts, and compared to an earlier study of head nods and syllable synchronization. The results indicated onset synchronization between head nods and the syllable in the short temporal domain and between the onset of longer gesture units and the talk spurt in a more extended temporal domain.

Introduction

There is currently considerable interest in the interaction between speech and gesture, and in particular the temporal relationship between prosody and gesture (Wagner et al., 2014). Kendon (1980) followed by McNeill (1992) have divided gestures into basic types differing in temporal scope. These are gesture units, gesture phrases and gesture phases. The gesture unit, the longest temporal domain, is the interval of gestural movement bounded by a period of non-movement. A gesture unit is comprised of one or more gesture phrases each of which can be divided up into a sequence of gesture phases. The stroke phase of a gesture phrase is particularly interesting in terms of prosody. Some of these strokes (also called "beat" gestures) often coincide and appear to be synchronized with prosodic and intonational peaks related to prominence such as pitch accents. For example, in studies by Leonard and Cummings (2011) and Loehr (2012), a correlation was found between the apices of strokes and focal accents in intonation. Flecha-Garcia (2007) studied alignment between eyebrow movement and pitch accents, and synchronization between three different phases of head nods and stressed syllables carrying focal accent was found by Alexanderson et al. (2013a, 2013b).

Synchronization between the phrase level of intonation and gesture phrases has also been studied by e.g. Karpinski et al. (2009) and Loehr (2012). Karpinski et al. (2009) found less synchronization between intonational phrases and gesture phrases, but they did note a high degree of overlap and a general "centering tendency" for semantically related gesture phrases and major intonational phrases. Loehr (2012) found a higher degree of synchronization between the gesture phrase and the intermediate intonational phrase as defined by Beckman and Pierrehumbert (1986).

Studying synchronization between speech and gesture has played an important role in building theories of human communication which approach speech and gesture production as arising from a common generation process (Kendon, 2004; McNeill, 2005). However, as seen above, most of the synchronization has been found in shorter domains of the accented syllable and the stroke gesture phase. In this study, we aim to investigate the relationship between longer domains of the gesture unit and talk spurts automatically extracted from spontaneous dialogue in order to test the hypothesis that there is not only a relationship between the intermediate phrase and gesture as detailed by Loehr (2012) but also between a longer period of speaker activation (the talk spurt) and the gesture unit. We further compare the timing results from our earlier studies of head nods and syllables with the timing relationship found between the gesture unit and the talk spurt. Ultimately we wish to explore the correspondences between co-speech, gestural domains and prosodic domains to try and more closely define the temporal domains of co-speech gestures.

Method

Corpus description

Portions of two dialogues taken from the Spontal corpus of Swedish dialogue were used for this investigation. The database, containing more than 60 hours of unrestricted conversation in over 120 dialogues between pairs of speakers, is comprised of synchronized high-quality audio and video recordings (high definition) and motion capture for body and head movements for all recordings with a frame rate of 100 frames per second (Edlund et al. 2010). During the recordings, the participants were seated in a sound studio and allowed to speak about any topic of their choice for 30 minutes. They remained in a seated position throughout the recording session. Figure 1 shows two video frames taken from the corpus. In this study we used a randomly selected five-minute passage from each of the two dialogues: Dialogue 1 between a male and a female participant, and Dialogue 2 between two male participants.

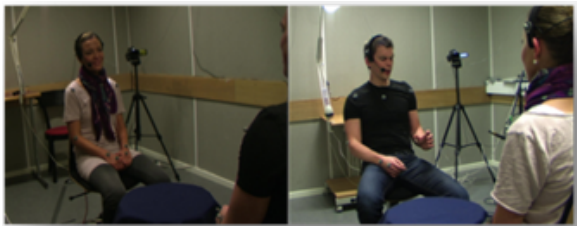

Figure 1. Frames taken from the spontal corpus of spontaneous dialogue

Hand movement detection

The motion data consist of the 3D positions of the motion capture markers attached to each subject. The total marker set contained 12 markers placed on the upper body and is illustrated in Figure 2.

Figure 2. Motion capture markers with 12 markers per subject.

Two different methods were used to quantify hand movements from the data and to divide the extracted movement into discrete sequences. Both methods are based on the conditions in the data corpus whereby the two participants are seated, facing each other, with their hands at a resting position on their laps. The first method (Method 1) is a simple, naïve method and interprets vertical displacement above a certain fixed threshold as gesture movement. Each gesture unit is defined as the time period from one resting position to the next.

The second method (Method 2) uses the velocity of the hand markers as the basis for segmentation. Typically, gesture units exhibit continuous hand motion except for short periods of time, when for example the hands change direction in a stroke. Our algorithm first detects periods of motion larger than 0.001 m/s and then merges segments with separation in time less than 200 ms. During a second pass, segments shorter than 200 ms are removed as well as segments which are performed in a horizontal plane (defined by a margin of 2.5 cm vertical distance). The last filter was introduced to remove non-communicative movement such as fidgeting.

Manual annotation

Two annotators independently annotated gesture phrases and phases using the ELAN annotation tool (Sloetjes and Wittenburg, 2008). The annotation procedure was carried out in two steps. The first step was to watch the dialogue at full speed and mark the beginning and end of any communicative hand gesture, one speaker at a time. The second step consisted of watching the gestures in slow motion and segmenting each gesture phrase into gesture phases following McNeill (1992) and Kita et al. (1998). The specified phases were the following: preparation, stroke, retraction, and pre- and post-stroke holds. All phases were optional except stroke which was an obligatory element of each marked gesture.

Automatic speech activity detection

To determine the speech activity at each frame, a voice activity detection algorithm (Laskowski et al., 2004) was applied to the audio recordings from the near microphones attached to the subjects. The speech activity was then divided into intervals of continuous speech, or talk spurts, following Brady (1968) and Heldner et

al. (2011). In this process, silent intervals of less than 200 ms are not regarded as silence but are integrated into the talk spurt. The process resulted in two sets of talk spurt segments for each dialogue, one for each subject. This procedure also captures short utterances such as humming and feedback signals which were not relevant to this study. Therefore, we introduced a pre-processing stage removing segments shorter than 500 ms.

Results

Evaluation of gesture extraction methods

To evaluate the automatic extraction methods we compared the results of the two methods with the manual annotations on a frame-by-frame basis for presence vs. absence of gestures for each dialogue. These results can then be compared to the average frame-by-frame agreement between the two annotators as an upper baseline (see Table 1). Accuracy is higher for method 2 than for method 1 for both dialogues. Therefore, results obtained from method 2 are used in the following analyses.

Table 1. Annotator agreement and automatic gesture extraction accuracy for the two dialogues and the two methods

	Annotator agreement	Accuracy: method 1	Accuracy: method 2
Dialogue 1	96%	80%	87%
Dialogue 2	99%	88%	96%

In cases where there were erroneous segments derived from method 2, a manual analysis of the characteristics of these segments was carried out. False positive gesture segments were in large caused mainly by pose shifts and fidgeting. Some undetected gestures (false negatives) were present when gestures were only performed by the fingers. Also long stroke holds were not detected as part of gestures as they do not exhibit any motion. However, as the motion capture data has a high frame rate, the method generally was more precise in detecting the exact onset and offset of the hand motion. This is of general importance for investigation of timing aspects and synchrony and can potentially be used to resolve inter-annotator non-agreement concerning the gesture segment boundaries.

Co-occurrence of gesture units and talk spurts

The number of talk spurts co-occurring with the automatically extracted gesture units varied considerably between the subjects and the dialogues. Dialogue 1 could be characterized as a dialogue rich in gesture with gesture units co-occurring in well over half of all the talk spurts, while dialogue 2 contained much fewer gesture sequences represented in only about 15% of the talk spurts. A general trend was that the talk spurts tended either to coincide temporally with a gesture unit throughout most of the duration of the talk spurt, or else the talk spurt contained no gestures at all.

Synchronization of talk spurts and gestures

The temporal relationships between the automatically extracted gesture units and the talk spurts are presented as box and whisker plots in Figure 3. The plots were calculated by measuring the timing difference between the onset of the talk spurt and the onset of the gesture unit. Positive values indicate that the talk spurt leads the gesture unit in time. There is considerable variation in the relationship between onset times, but there is a central tendency in which the onset of the talk spurt slightly precedes the onset of the gesture unit. Greater variability is seen for the two speakers with the most gestures in Dialogue 1.

Synchronization of head nods and syllables

In an earlier study (Alexanderson et al., 2013a) temporal synchronization was studied between head nods annotated as having a beat function and anchor points of the syllable. Both the head nods and the syllable anchor points were automatically extracted. Figure 4 shows the time difference between two different anchor-points of the syllable (onset and nucleus) and three different phases of the nod: peak velocity of the downward phase (p1), max rotation (p2) and peak velocity of the upward phase (p3) for one speaker from the spontal corpus. The timing relationship between the gesture and the syllable does not seem to be influenced by the choice of syllable anchor-point. On an average the nod begins slightly before the syllable onset with the maximum rotation of the nod centering on the syllable nucleus.

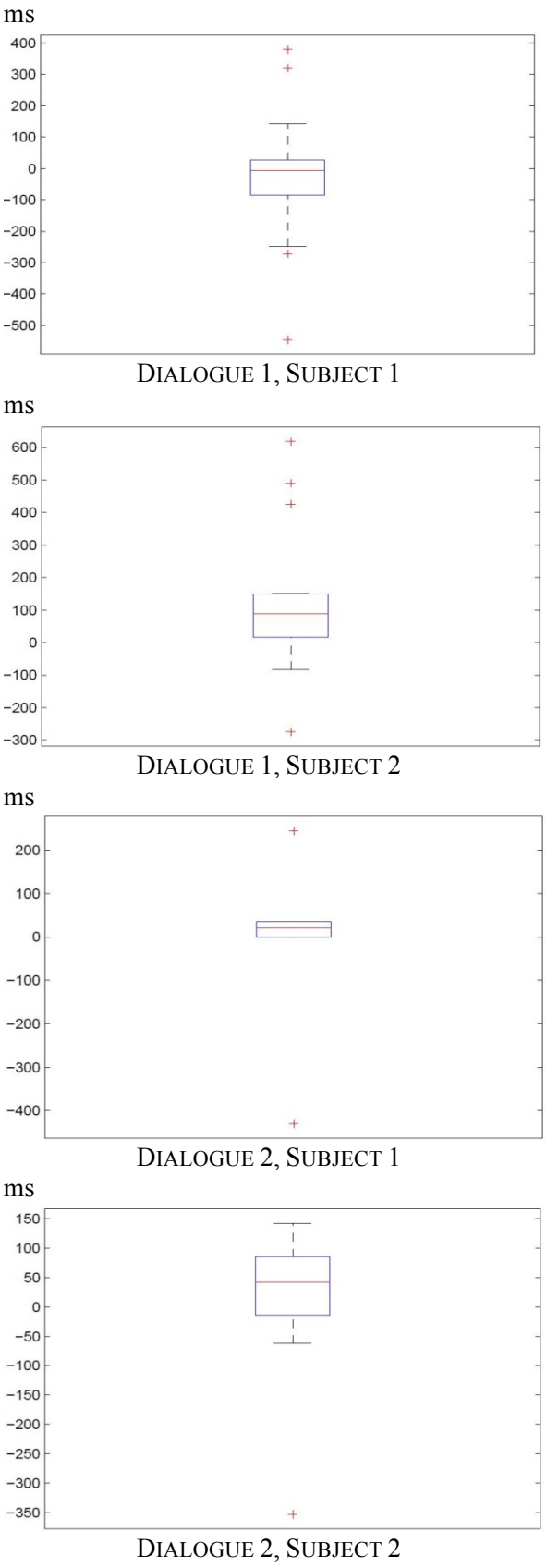

Figure 3. Box and whisker plots showing the temporal relationship between gesture unit and talk spurt. Positive values indicate that the talk spurt precedes the gesture unit.

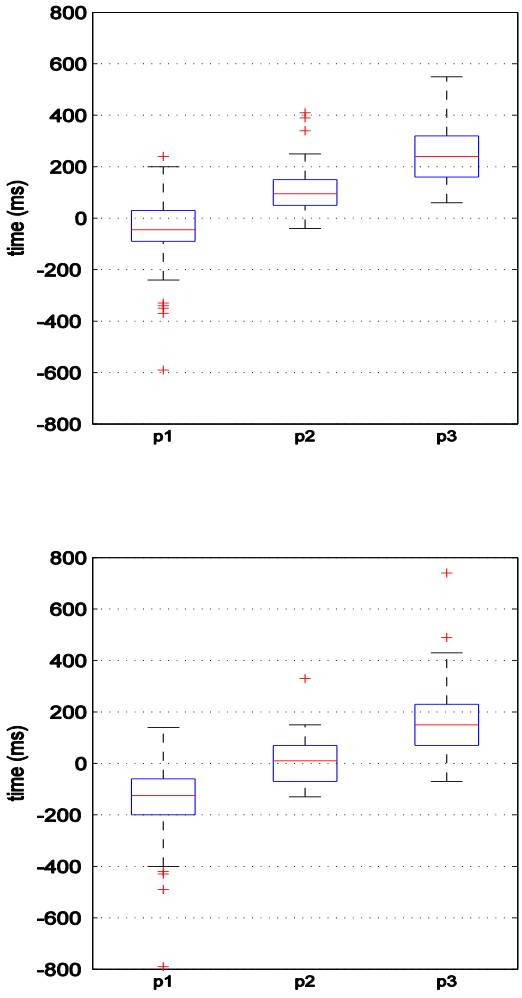

Figure 4. Timing of syllable anchor-points, onset (top) and nucleus (bottom), with respect to three different phases of the head nod: peak velocity of the downward phase (p1), max rotation (p2) and peak velocity of the upward phase (p3). Negative values indicate that the head nod phase precedes the syllable anchor point (from Alexanderson et al. 2013a).

Discussion

In this study we have explored the use of automatic methods to locate co-speech gestures and the co-occurring speech in longer sequences of spontaneous dialogue. We have investigated the temporal aspects of these sequences and seen a co-occurrence of what has in the literature been termed gesture units with automatically extracted talk spurts. This study was limited to two dialogues, but even in only two dialogues, we see a large variation in the incidence of gestures occurring during the talk spurts. One of the dialogues was rich in gesture with over half

of the talk spurts co-occurring with gesture. The other dialogue showed much less gesture activity for both participants. This points to the optionality of producing gesture units which temporally coincide with a complete talk spurt. However, in the gesture-rich dialogue, the gesture units were often found to coincide with the talk spurt and together formed a coordinated speech-gesture unit.

The use of an automatic analysis of the temporal synchronization of gesture movement and talk spurts can help us in our understanding of the domain of co-occurring speech and gesture. The results obtained here are consistent with the findings by Loehr (2012) and Karpinski et al. (2009) in that there is less absolute synchrony between gesture and speech on the phrase level than on the syllable level. The gesture sequences extracted by the automatic gesture movement analysis presented here most closely correspond to the gesture units rather than gesture phrases. The synchronization between the gesture units and the talk spurts shown in Figure 3 displays considerable variation, but also shows a central tendency where the onset of gesture tends to be coordinated with the onset of the talk spurt with the talk spurt slightly preceding the onset of the gesture unit on an average. The initiation phase of the talk spurt may be preceded by gaze activity and posture shifts, but hand movement seems be initiated slightly after the beginning of speech.

This constitutes a timing trend contrary to that appearing between head motion and the syllable as shown in Figure 4. Beat gestures can thereby be seen to share the time domain of the syllable while gesture units share the time domain of the talk spurt. Moreover, this could indicate that on a global temporal domain, speech precedes gesture, while on the local domain of the syllable where beat gestures and accented syllables co-occur, gesture precedes the prosodic correlate having the same function.

Finally, the analysis of gesture and speech in a longer temporal domain points to the possibility of defining the talk spurt as analyzed here as a speech correlate to the gesture unit. This domain is longer than the intonational phrase and can be seen as bodily activation in both speech and gesture comprising an important temporal domain in spontaneous dialogue.

Acknowledgements

The work reported here has been funded by the Bank of Sweden Tercentenary Foundation (P12-0634:1) and the Swedish Research Council (VR 2010-4646). We would also like to thank Meg Zellers for help in preparing and executing the gesture annotation and to Jens Edlund for assistance with the Spontal corpus.

References

Alexanderson S, House D and Beskow J (2013a). Extracting and analyzing head movements accompanying spontaneous dialogue. In *Proc. Tilburg Gesture Research Meeting.* Tilburg University, The Netherlands.

Alexanderson S, House D and Beskow J (2013b). Aspects of co-occurring syllables and head nods in spontaneous dialogue. In *Proc. of 12th International Conference on Auditory-Visual Speech Processing (AVSP2013)*. Annecy, France.

Beckman M and Pierrehumbert J (1986). Intonational structure in Japanese and English. *Phonology Yearbook III*: 15-70.

Brady P T (1968). A statistical analysis of on-off patterns in 16 conversations. *The Bell System Technical Journal,* 47: 73-91.

Edlund J, Beskow J, Elenius K, Hellmer K, Strömbergsson S and House D (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. In Calzolari N, Choukri K, Maegaard B, Mariani J, Odjik J, Piperidis S, Rosner M and Tapias D, eds, *Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valetta, Malta, 2992-2995.

Flecha-Garcia M L (2007). Non-verbal communication in dialogue: Alignment between eyebrow raises and pitch accents in English. In *Proceedings of CogSci-2007*. Austin, Texas, USA, 1753.

Heldner M, Edlund J, Hjalmarsson A and Laskowski K (2011). Very short utterances and timing in turn-taking. In *Proceedings of Interspeech 2011*. Florence, Italy, 2837-2840.

Karpinskyi M, Jarmolowicz-Nowikow E and Malisz Z (2009). Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues. *Speech and language technology,* 11: 113-122.

Kendon A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In Key M R ed, *The relationship of verbal and nonverbal communication*. The Hague: Mouton, 207-227.

Kendon A (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita S, van Gijn I and van der Hulst H (1998). Gesture and Sign Language in Human-Computer Interaction. In Wachsmuth I and Frölich M eds, *Lecture Notes in Computer Science* 1371: Springer, 23-35.

Laskowski K, Jin Q and Schultz T (2004). Crosscorrelation-based multispeaker speech activity detection. Proceedings of *Interspeech 2004*. Jeju Island, South Korea.

Leonard T and Cummins F (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes,* 26: 1457−1471.

Loehr D (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. Laboratory Phonology. *Journal of the Association for Laboratory Phonology*, 3: 71-889.

McNeill D (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.

McNeill D (2005). *Gesture and thought.* Chicago: University of Chicago Press.

Sloetjes H and Wittenburg P (2008). Annotation by category – ELAN and ISO DCR. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Wagner P, Malisz Z and Kopp S (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57: 209-232.

Intonations and functions of questions in Helsinki Swedish conversations

Martina Huhtamäki
Finnish, Finno-Ugric and Scandinavian studies, University of Helsinki, Helsinki, Finland

Abstract

This is a qualitative study of questions in the Swedish variety spoken in Helsinki, Finland. Focus is on the intonation contours and functions of the questions. The data consist of recordings of spontaneous conversations. The study connects to the methodological framework of interactional linguistics, and the questions are analyzed sequentially and phonetically.

The results show that the intonation contours of the questions resemble those of Sweden Swedish and Finnish. There is no direct relationship between the intonation contour of a question and its function, but intonation is used to distinguish between utterances inside the category of questions. Also other features take part in the function of the question, that is, syntactical, lexical, sequential and epistemic factors.

Introduction

Questioning is an essential activity in conversation, which has various functions (Hayano, 2013; Stivers & Enfield, 2010). It is assumed that intonation contributes to the signaling of questions universally, mainly through final rising intonation (e.g. Gussenhoven, 2004). In many languages, especially yes/no-questions have final rising intonation, whereas wh-questions have falling intonation (e.g. Bolinger, 1989 on English). However, Couper-Kuhlen (2012) has shown for English that not only syntax, but also action-type and epistemic factors affect the final intonation of a question. Thus the relationship between intonation and function in questions is a complex one.

Helsinki Swedish is an interesting case in point, as it syntactically and lexically is close to Sweden Swedish (Reuter, 2006; Wide & Lyngfelt, 2009), but prosodically and phonetically resembles Finnish (Aho 2010; Kuronen & Leinonen 2008). For example, Helsinki Swedish lacks tonal accents, like most Finland Swedish varieties (Bruce, 2010; Selenius, 1974).

Thus, the most relevant languages to compare Helsinki Swedish with are Sweden Swedish and Finnish. According to Strömbergsson, Edlund and House (2012), spontaneous questions in Sweden Swedish dialogues vary regarding several prosodic features, that is, final intonation, pitch variation and duration. Final rising intonation is mainly found in backward-looking wh-questions, like *what?* and *what did you say?*. House (2005) regards rising intonation as an optional interrogative feature in spontaneous wh-questions, functioning response-seeking and expressing a friendly attitude. In Finland Swedish wh-questions, Kuronen and Leinonen (2010) have found a falling contour in, starting with a rise-fall on the question-word and having a smaller pitch-movement on the nominal element at the end.

For Finnish, Iivonen (1978) presents six intonation contours that occur in spontaneous questions: 1) falling, 2) high initial, 3) extra high initial, 4) high overall until the last stressed syllable, 5) rising from beginning and 6) final rising. Anttila (2008) includes four more contours: a rising-falling, a level, a falling-rising, and a globally low contour. Anttila points out that creak is more common in the questions than in the statements in her data. According to Iivonen and Anttila, the intonation contour of a question is related to syntax, discourse function and the speaker's idiolect. A globally rising contour is according to Iivonen used to indicate astonishment and call for repetition. The final rising contour indicates according to Iivonen that floor is open, and may be a result of foreign influence and an idiolectical feature.

Data

The research data consist of 6 recordings of spontaneous conversations. In the conversations 29 persons in all take part, aged 9 to about 60 years, both females and males. The participants are Swedish speakers from the Helsinki region in Finland. The conversations can be characterized as everyday conversations. They are multiperson conversations, as three or more persons take part in each conversation.

Questions are extracted from the recordings according to two criteria. Firstly, they concern the epistemic domain of the recipient. This means that the speaker has less knowledge than the recipient about the topic of the question. Therefore, rhetorical questions are not included, as the speaker then has more knowledge than the recipient. Secondly, questions are utterances that make an answer necessary. A consequence of the definition is that it includes utterances with interrogative and non-interrogative syntax.

Methods

The theoretical and methodological framework is Interactional Linguistics (IL) (Couper-Kuhlen & Selting, 2001). In IL, researchers study how linguistic features, like intonation, are used to create meanings in interaction. In IL, methods from Conversation Analysis (CA) are combined with other methods from linguistics, for example phonetic analysis.

In this study I have performed a sequential analysis of the sequences where the questions appear, as well as a phonetic analysis of the intonation of the questions. The sequential analysis includes an analysis of the syntactical and lexical features of the question, its sequential placement as well as of situational factors. The sequential analysis results in an account of the function of the question. The intonation of the question is studied with auditory and acoustic methods. For the acoustic analysis I have used the program Praat (Boersma & Weenink, 2015). As part of the analysis, I have transcribed the sequences with the questions (cf. Transcription symbols). To get an impression of where in the speaker's pitch range a question is produced, I have measured the modal pitch range of each speaker on 1-2 minutes of speech. In the acoustic records the maximum, minimum and median of the pitch range are presented as horizontal lines.

Results and discussion

The study shows that no single intonation contour can be perceived as indicating questioning in Helsinki Swedish, but questions are produced with various intonation contours. Most of the contours have final falling intonation (ca 80 %). The part of questions with final rising intonation is about the same as House (2005) found in Sweden Swedish wh-questions (under 20 %). In addition, a small part of the questions have final level intonation.

The intonation contours of the questions resemble those described by Iivonen (1978) and Anttila (2008) for Finnish. Furthermore, creak is a common turn-final feature of the questions (cf. Anttila, 2008). Questions with final rising or level intonation have similar functions in the conversations as Strömbergsson (2012) have described for Sweden Swedish. Consequently, intonation in Helsinki Swedish questions is used in ways that both resemble Finnish and Sweden Swedish.

In Example 1 (line 2), a yes/no-question is used to request information about whether the other participants do watch a certain TV-series. The question gets two latched answers in the affirmative (lines 3, 4). The question introduces a new sub-topic inside the current topic "TV-series".

Example 1[1]*. Fat man (Sewing Circle)*

01 M: ne utan di åker omkring där å de e
no but they drive around there and it is

vackra scenerier å sånt.
beautiful scenarios and such

02 A: **nåmen brukar ni titta på Hund begraven.=**
but do you regularly watch Jake and the Fat man

03 T: =[[jå:å?
ye:es

04 E: =[[jå?
yes

The intonation contour of the question is level until the focal accent, which is produced as a pitch peak (cf. fig. 1). The final intonation is falling. The intonation contour resembles type 4) in Iivonen (1978), but the level stretch is not so high. Questions with this pitch patterns are regularly used to introduce a new topic in the conversation. Hence, the crucial feature for the

[1] Analyzed in Huhtamäki (2012).

choice of intonation contour is not only the syntax of a question, but its function. There is not one intonation contour used for yes/no-questions and one used for wh-questions in the data. The contour described for wh-questions in Finland Swedish by Kuronen and Leinonen (2010) is not frequent in the data, not in wh-questions, nor in other types of questions.

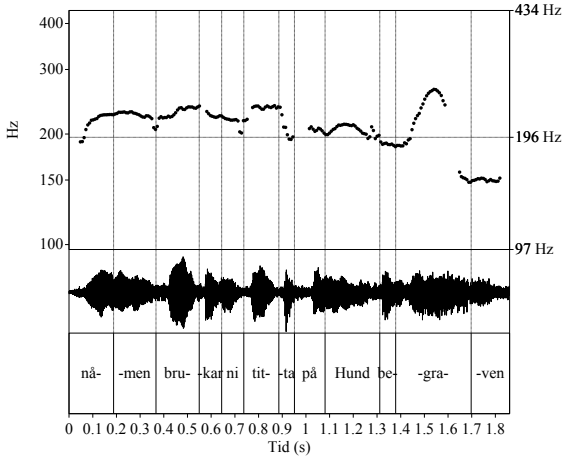

Figure 1. Pitch trace and waveform of the question "nåmen brukar ni titta på Hund begraven" (female speaker).

Furthermore, questions have many functions in the studied conversations, ranging from seeking information, seeking confirmation, initiating repair, introducing a new topic, mobilizing a response and expressing an affective stance (cf. Drew & Couper-Kuhlen, 2014; Halonen & Sorjonen, 2008; Heritage, 2012; Labov & Fanshel, 1977; Schegloff, Jefferson, & Sacks, 1977).

There is no direct relationship between the intonation contour of a question and its function. Instead intonation is used to distinguish between utterances inside the category of questions. Intonation contour does rarely contribute to the function alone, but together with syntactic and lexical features of the question, as well as its sequential placement and the epistemic relations between the participants.

Final rising or level intonation is for instance used to signal a trouble in a previous turn, that is, for repair initiation (cf. Anttila, 2008; Iivonen, 1978 on call for repetition in Finnish; cf. Strömbergsson et al., 2012 on backward-looking *vad* in Sweden Swedish). However, also other features in the utterance and the sequence may contribute to this function, for example the use of the question-word *va* 'what', repetition of an element of the previous turn, and final discourse markers, like *då* 'then'.

In Example 2 (line 2), the question-word *va* together with the initial discourse marker *aj* 'oh', and the final discourse marker *då* 'then' are used to initiate repair on a previous turn (line 1). One repair-solution is performed in overlap with the repair-initiation (line 3), and another repair-solution after that turn (line 4). Both repair-solutions give more information about whom Johanna is talking, treating the trouble as being about an underspecified referent (cf. Egbert et al., 2009).

Example 2. Not me (College Language)

01 J: de: int jag.
it's not me

02 S: **aj va[då?**
PRT what PRT

03 A: [de: Mia.
it's Mia

04 J: Mia å Sandra;
Mia and Sandra

The intonation of the question is globally falling with a turn-final rise over a small pitch span (fig. 2). This contour resembles type 6 in Iivonen (1978).

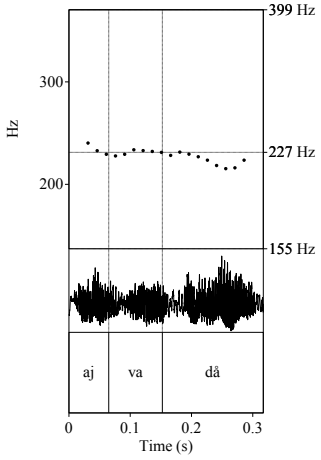

Figure 2. Pitch trace and waveform of the repair initiator "aj vadå" (young female speaker).

Similarly, an affective stance may be shown with a question that has a wide pitch span and possibly final rising intonation over a wide pitch span (cf. Iivonen, 1978 on astonishment in Finnish). Also here, other features contribute to the function, like some piece of surprising news or a previous utterance expressing a stance, the

verbal form of the question (e.g. *varför* 'why'), as well as the repetition of an element of the previous utterance.

In Example 3 (line 3) a noun-phrase is used for checking the information of a previous turn and showing an affective stance towards it. A group of young people are discussing that one student failed the music class in school. Jocke displays his disbelief against that fact, and continues in his later turn to explicate his stance (line 6).

Example 3. Failed in music (Summer Camp)

01 S: å nån fick ju (.) nån fick ju underkänt också i de.
and somebody failed (.) somebody failed in it

02 Ja: f [ick den. (.) vem]
did it (.) who

03 Jc: [**i MUSSA.**]
in music

04 M: [[Eva Alén säkert,
Eva Alén certainly

05 Ja: [[vem. ((skratt))
who ((laughter))

06 Jc: på riktit man måst va en rikti kråka fö de,
for real you really have to be a crow to do that

The question is characterized by great pitch movements (fig. 3). Pitch rises on the stressed syllable over a wide pitch span towards the top of the speaker's pitch range and falls at the end. This type of contour is described by Anttila (2008) as rising-falling.

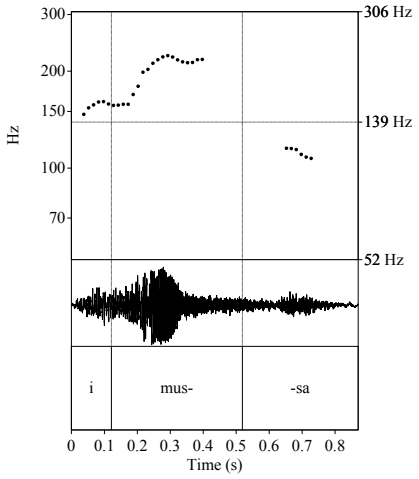

Figure 3. Pitch trace and waveform of the stance-taking question "i mussa" (young male speaker).

The examples above show some of the functions and intonation contours of the questions in the Helsinki Swedish data. There are also other contours used in different functions in the data. The examples demonstrate how several features take part in the function of a question, of which intonation contour is but one feature.

Conclusion

The results display that various intonation contours are used in questions in Helsinki Swedish. Also functionally, the questions form a heterogeneous group.

The intonation contour contributes to the function of the questions in a specific context. Intonation contour is together with other features in that context used for designing questions with specific functions. Hence, intonation contour is not a means to signal question-mode as a sentence-type. Instead, it works as a contextualization cue (Gumperz, 1982). By this, I mean that an intonation contour get its meaning and contributes to the meaning of a question in its context. Therefore, the meanings of intonation contours are not intrinsic, but context-dependent.

The results also support the conception that Helsinki Swedish intonation resembles both Sweden Swedish and Finnish intonation. As for those languages, falling intonation dominates in the data. Intonation contours with final rising and level intonation are shown to have similar functions as in these languages. Further comparisons between Finland Swedish, Swedish Swedish and Finnish are welcome.

Acknowledgements

I acknowledge The Ella och Georg Ehrnroot foundation, Svenska kulturfonden and Svenska litteratursällskapet for financial support. I also thank Jan Lindström and Anne-Marie Londen for useful comments.

Transcription symbols

.	final falling intonation
?	final rising intonation
,	final level intonation
syllable	stressed syllable
[word]	start and end of overlap
[[word	simultaneously beginning turns
(.)	short pause
(0.5)	measured pause
(( ))	transcriber's comments

References

Aho E (2010). *Spontaanin puheen prosodinen jaksottelu*. Pic Monographs. Helsinki: Nykykielten laitos, yleinen kielitiede, humanistinen tiedekunta, Helsingin yliopisto.

Anttila H (2008). *The effect of interrogative function on intonation in spontaneous and read Finnish*. Helsinki: Department of speech sciences, University of Helsinki.

Boersma P & Weenink D (2015). Praat: doing phonetics by computer. Retrieved from www.fon.hum.uva.nl/praat/

Bolinger D (1989). *Intonation and its uses. Melody in grammar and discourse.* London: Edward Arnold.

Bruce G (2010). *Vår fonetiska geografi*. Lund: Studentlitteratur.

Couper-Kuhlen E (2012). Some truths and untruths about final intonation in conversational questions. In J P de Ruiter, ed., *Questions. Formal, functional and interactional perspectives.* Cambridge: Cambridge University press, 123–145.

Couper-Kuhlen E & Selting M (2001). Introducing interactional linguistics. In M Selting & E Couper-Kuhlen, eds., *Studies in interactional linguistics.* Amsterdam/Philadelphia: John Benjamins publishing company, 1–22.

Drew P & Couper-Kuhlen E (2014). Requesting - from speech act to recruitment. In P Drew & E Couper-Kuhlen, eds., *Requesting in social action*. Amsterdam: John Benjamins publishing company, 1–34.

Egbert M, Golato A & Robinson J D (2009). Repairing reference. In J Sidnell, ed., *Conversation analysis: comparative perspectives*. Cambridge: Cambridge University press, 104–132.

Gumperz J J (1982). *Discourse strategies*. Cambridge: Cambridge University press.

Gussenhoven C (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University press.

Halonen M & Sorjonen M-L (2008). Using *niin*-interrogative to treat the prior speaker's action as an exaggeration. *Discourse studies*, *10*(1): 37–53.

Hayano K (2013). Question design in conversation. In J Sidnell & T Stivers, eds., *The handbook of conversation analysis*. Chichester: Wiley-Blackwell, 395–414.

Heritage J (2012). Epistemics in action: action formation and territories of knowledge. *Research on language & social interaction*, *45*(1): 1–29.

House D (2005). Phrase-final rises as a prosodic feature in wh-questions in Swedish human–machine dialogue. *Speech Communication*, *46*(3-4): 268–283.

Huhtamäki M (2012). Prosodiska mönster hos frågor. En undersökning av Helsingforssvenska samtal. *Språk och stil 22*(2): 153–184.

Iivonen A (1978). Is there interrogative intonation in Finnish? In E Gårding, G Bruce, & R Bannert, eds., *Nordic prosody. Papers from a symposium*. Lund: Department of linguistics, Lund university, 43–53.

Kuronen M & Leinonen K (2008). Prosodiska särdrag i finlandssvenska. In M Nordman, ed., *Svenskans beskrivning 29, Vasa universitet.* Skrifter utgivna av Svensk-Österbottniska samfundet r.f. nr 70. Vasa: Vasa universitet, 161–169.

Kuronen M & Leinonen K (2010). Prosodin i frågeordsfrågor i finlandssvenskan. In C Falk, A Nord, & R Palm, eds., *Svenskans beskrivning 30. Förhandlingar vid trettionde sammankomsten för svenskans beskrivning. Stockholm den 10 och 11 oktober 2008*. Stockholm: Institutionen för nordiska språk, Stockholms universitet, 165–176.

Labov W & Fanshel D (1977). *Therapeutic discourse : psychotherapy as conversation*. New York: Academic Press.

Reuter M (2006). Svenskan i Finland på 1900-talet. In A-M Ivars, M Reuter, P Westerberg, & U Ådahl-Sundgren, eds., *Vårt bästa arv. Festskrift till Marika Tandefelt den 21 december 2006.* Skrifter utgivna vid Svenska handelshögskolan nr 165. Helsingfors: Svenska handelshögskolan, 29–45.

Schegloff E A, Jefferson G & Sacks H (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2): 361–382.

Selenius E (1974). *Helsingforssvensk ettordsaccentuering*. Helsinki: Helsingin yliopiston fonetiikan laitoksen julkaisuja, Helsingin yliopisto.

Stivers T & Enfield N J (2010). A coding scheme for question–response sequences in conversation. *Journal of pragmatics*, *42*(10): 2620–2626.

Strömbergsson S, Edlund J & House D (2012). Question types and some prosodic correlates in the Spontal corpus of Swedish dialogues. In A Eriksson & Å Abelin, eds., *Proceedings Fonetik 2012, the XXVth Swedish phonetics conference, May 30-June 1, 2012*. Gothenburg: Deparment of philosophy, linguistics and theory of science, University of Gothenburg, 49–52.

Wide C & Lyngfelt B (2009). Svenskan i Finland, grammatiken och konstruktionerna. In C Wide & B Lyngfelt (Eds.), *Konstruktioner i finlandssvensk syntax: skriftspråk, samtal och dialekter*. Helsingfors: Skrifter utgivna av Svenska litteratursällskapet i Finland, 11–43.

A pilot study: acoustic and articulatory data on tonal alignment in Swedish word accents

Malin Svensson Lundmark, Johan Frid and Susanne Schötz
Centre for Languages and Literature, Lund University

Abstract

This pilot study compares the timing of articulatory gestures to the timing of the tonal contour in South Swedish Accent 1 and Accent 2. Acoustic and articulatory data were collected with an EMA (Carstens AG501). Variables included the tonal alignment of the high tone H and the following low tone L to the vowel onset, the syllable offset, as well as to the lip aperture and the tongue body. Acoustic results point towards different units as host for the accents: Accent 1 aligns with the vowel while Accent 2 aligns with the syllable. The articulatory data shows alignment to different gestures: a stable tonal alignment with the lip aperture in Accent 1, and a less stable alignment with the movement of the tongue body in Accent 2.

Introduction

In the prosodic typology of Swedish intonation provided by the Lund Model (Bruce & Gårding, 1978; Bruce, 2007) the two word accents are assumed to be represented by a tonal fall associated with the stressed syllable. However, the tonal peak of Accent 1 always precedes the peak of Accent 2 in all tonal dialect types of Swedish. Moreover, there are extensive timing differences between dialects, e.g. Accent 1 by a speaker of Stockholm Swedish begins with a low tone, while in South Swedish it starts with a high tone.

There are morphological rules attached to the Swedish word accent distinction (see e.g. Bruce, 1998; Riad, 2014; Riad 2012). For example, a nominal monosyllabic word stem is assigned Accent 1 in the singular form, but receives Accent 2 when plural suffixes are added. Perception studies have found evidence that word accents indeed provide cues of the upcoming suffix (Roll et al., 2013).

However, views on the phonological typology differ: the Lund Model assesses an equipollent distinction where both accents have lexical tones. Other accounts have stressed a privative distinction where only Accent 2 is lexically marked (Riad, 2006; Engstrand, 1997). It has also been suggested that for some dialects the lexical tone in Accent 2 consists of a high tone, while in some dialects (including South Swedish) it is a low tone (Riad, 2006). In the revised Lund Model, Bruce (2007) made the specific assumption for South Swedish of a tonal fall, an H+L pattern, for Accent 1 and a rise, an L+H pattern, for Accent 2. The rise in South Swedish Accent 2 has indeed been shown to be relevant from a perceptual point of view (Ambrazaitis & Bruce, 2006). However, in this pilot study we will only look at the stability of H and L of the fall in both accents, and not L of the rise.

Studies in intonational languages have displayed an unambiguous case of the start of the rise L aligning with the syllable onset in pre-nuclear accents, e.g. Greek (Arvaniti et al., 1998), Italian (Niemann et al., 2011), Dutch (Caspers & Van Heuven, 1993), English (Ladd et al., 1999), and German (Atterer & Ladd, 2004), or just after the syllable onset in the tone language Mandarin (Xu, 1998). A similar consistent result has not been found for the high target H in either of the studies. However, in a study on tonal alignment in the South Swedish Accent 2, the L marking the beginning of the rise appeared to be less stable than H (Svensson Lundmark, 2014).

Recent tonal alignment studies have incorporated articulatory data (Hermes et al., 2008; Mücke et al., 2012; Niemann et al., 2014), following the articulatory phonology framework and the notion of articulatory gestures (Browman & Goldstein, 1992). By including intonation in the gestures, i.e. a tonal gesture, it is possible to couple it with the consonantal and vocalic gestures (Mücke et al., 2012; Niemann et al., 2011). Niemann et al. (2014) found a stable anchoring of H at the vocalic gesture in rising nuclear accents in German, hence presented evidence for stable L as well as H targets.

In this pilot study we adopt this promising account for the study of the Swedish word accents. This enables us to address a number of important questions: If the timing of the tonal curve equals a tonal gesture, are the tonal gestures of Accent 1 and Accent 2 coupled with separate articulatory gestures? Or do the articulatory gestures differ between the accents? Any such difference would open up for the possibility that the two word accents have different roles in speech motor control. Maybe this would shed a light on the issue of whether the word accents should be considered a privative or an equipollent distinction. To the best of our knowledge articulatory studies in the past on the Swedish word accents have been restricted to laryngeal control (e.g. Gårding et al., 1975).

Method

Speech material

The material consisted of simplex disyllabic Accent 1 and Accent 2 target words with stress on the penult produced in the carrier phrase [Det var TARGET jag sa.] (It was TARGET I said.). In the sentence context the target words elicited a focal accent, which in the South Swedish dialect does not differ distinctively from a non-focal accent [8]. The words were matched so that each Accent 1 – Accent 2 pair consisted of the same nominal word stem, but with different suffixes: definite singular for Accent 1 and indefinite plural for Accent 2. The target words also conditioned either an open or a closed stressed syllable (CV:.CVC or CVC.CVC), which in turn contained either a closed or an open vowel (Table 1). Thus, the material consisted of eight target words; each accent pair conditioned syllable type and vowel type.

Data collection

Two female speakers of South Swedish (age 38 and 49) read the material ten times each, i.e. 80 target words per speaker. The sentences were shown on a prompter in a random order.

Sound recordings and kinematic data were recorded simultaneously using a 3D Electromagnetic Articulograph (Carstens AG501) with an external condenser microphone (t.bone EM 9600). Articulatory movements were tracked by sensors on the upper and lower lips (at the vermilion border in the sagittal plane), on the tongue body and also on the bridge of the nose and behind one ear; the latter two sensors were used to correct for head movements during the recordings.

Table 1. The conditions of the eight target words.

		Closed vowel	Open vowel
Accent 1	CVC	/bilden/ (the picture)	/valen/ (the mound)
	CV:	/biːlen/ (the car)	/vɑːlen/ (the whale)
Accent 2	CVC	/bilder/ (pictures)	/valar/ (mounds)
	CV:	/biːlar/ (cars)	/vɑːlar/ (whales)

Measurements

The recorded samples amounted to 160 tokens (2 speakers x 8 target words x 10 repetitions). Acoustic segmentation and annotation of F0 turning points was made manually in Praat (Boersma & Weenink, 2014). For the high tone (H) maximum pitch was labelled in the tonal peak or high plateau. For the following low point (L2) the minimum F0 was used unless an apparent turning point appeared later or earlier.

Speaker MS occasionally used atypical South Swedish accent patterns. These samples were omitted. Some target words were produced with creaky voice, which resulted in some missing data for the L2 point. Out of 160 observations 144 with H-targets and 126 with L2 were used in our further analyses. The data of both speakers were collapsed in the analysis. From the acoustic data the following variables were obtained (Figure 1):

1) the distance from H to the vowel onset
2) the distance from L2 to the vowel onset
3) the distance from H to the syllable offset
4) the distance from L2 to the syllable offset

The two articulatory targets were automatically annotated in R (R Core Team, 2015) (marked as ‘x’ in Figure 1) and the following variables were collected:

5) the distance from H to the max velocity of the lip aperture
6) the distance from L2 to the max velocity of the lip aperture
7) the distance from H to the articulatory target of the vocalic gesture (max tongue body for open vowel, min for closed vowel)
8) the distance from L2 to the articulatory target of the vocalic gesture (max tongue body for open vowel, min for closed vowel)

9) the sync difference (distance between the two articulatory targets)

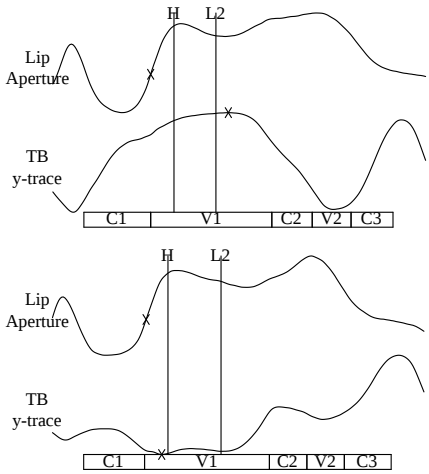

Figure 1. Acoustic and articulatory landmarks for the tonal targets H and L2. Accent 1 words by speaker SS: open syllables (CV:.CVC) with a closed vowel (top) and an open vowel (bottom).

Results

Acoustic results

Duration

The duration was similar for Accent 1 and Accent 2 in the different target words, stressed syllables or vowels (Table 2). We also found similar durations in the syllable types (CVC and CV:), except for the obvious difference in vowel duration. Also, a significant difference was found between the two stressed vowel types closed and open (SE=11.39, t=-2.421, p<.05).

Table 2. Duration (ms) of the variables used in the study.

Predictor variables	Target words	Stressed syllable	Stressed vowel
Accent 1	724	433	209
Accent 2	722	426	212
CVC	721	426	143
CV:	725	433	276
Closed vowel	732	442	197
Open vowel	714	416	224

Acoustic tonal alignment

In Accent 1 both tonal targets seem stable in relation to the vowel onset: H is about 50 ms after the vowel onset and L2 about 200 ms (see Figure 2). In Accent 2 H is about 240 ms and L2 about 400 ms after the vowel onset, and also more variable than in Accent 1. A regression analysis shows that the syllable type only affects the alignment of H in Accent 2 (SE=0.007, t=-4.209, p<.001). However, the target word /bilder/, which has a stressed CVC syllable with a closed vowel, appears to stand out. When we removed the target word /bilder/ from the data, the syllable type also influenced L2 in Accent 2 (SE=0.031, t=-2.79, p<.01). The effect of the syllable type indicates that Accent 2 is not aligned to the vowel.

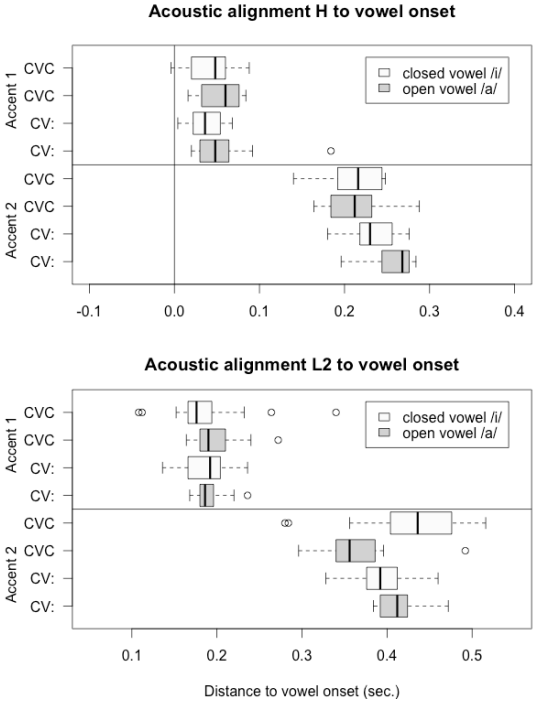

Figure 2. Alignment of H (top) and L2 (bottom) to the vowel onset (vertical line at 0 sec.).

Both targets seem to vary in their alignment to the syllable offset (see Figure 3) in Accent 1, while in Accent 2, H is quite stable about 140 ms before the syllable offset and L2 is aligned almost perfectly to the syllable offset. According to a regression analysis the syllable type affects the timing of H and L2 significantly in both Accent 1 (H: SE=0.009, t=5.576, p<.001; L2: SE=0.011, t=5.637, p<.001) and in Accent 2 (H: SE=0.006, t=3.936, p<.001; L2: SE=0.012, t=5.664, p<.001). However, removing the target word /bilder/, which in Figure 3 clearly deviates from the others, the significant effect disappears in Accent 2 for H (SE=0.004, t=-1.475, p=0.146) and for L2 (SE=0.029, t=-1.196, p=0.237). This indicates a stable alignment with the syllable in Accent 2.

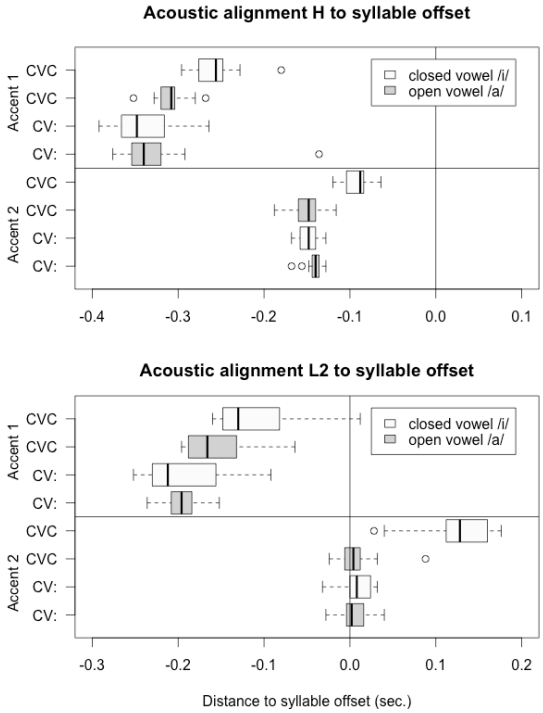

Figure 3. Alignment of H (top) and L2 (bottom) to the syllable offset (vertical line at 0 sec.).

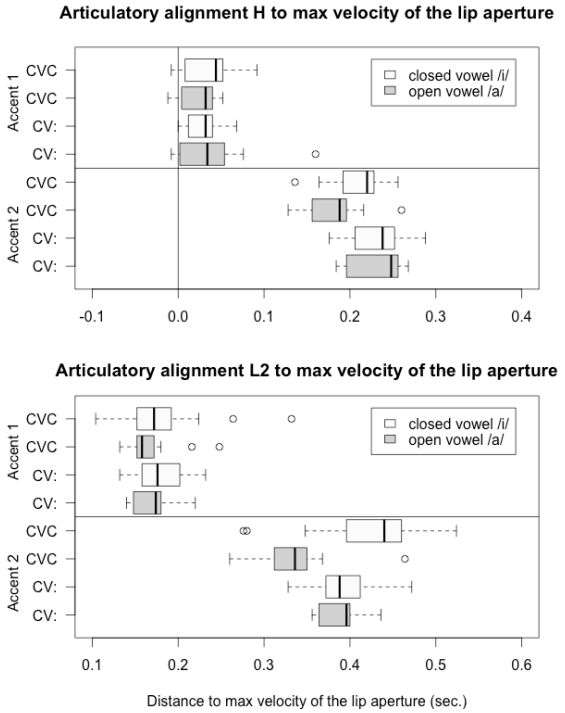

Figure 4. Alignment of H (top) and L2 (bottom) to max velocity of the lip aperture (vertical line at 0 sec.).

Articulatory tonal alignment

The tonal alignment differs significantly between the accents in relation to the max velocity of the lip aperture in H (SE=0.006, t=-33.63, p<.001) as well as in L2 (SE=0.008, t=-26.03, p<.001). Accent 1 displays a more stable relationship to the lip aperture than Accent 2 (see Figure 4). A significant effect of syllable type is only found in Accent 2 for H (SE=0.008, t=-4.581, p<.001). By excluding /bilder/ once again a significant effect in Accent 2 is also found for L2 (SE=0.031, t=-3.188, p<.01), indicating an alignment to the lip aperture in Accent 1.

In Figure 5 the target of the vocalic gesture displays less stability in the tonal alignment than in the lip aperture, but a significant difference is still found between the accents in both H (SE=0.011, t=-17.64, p<.001) and L2 (SE=0.013, t=-17.80, p<.001). The syllable type affects Accent 1 in both H (SE=0.016, t=4.211, p<.001) and L2 (SE=0.016, t=4.368, p<.001). Excluding /bilder/ results in no significant effect of syllable type on neither H nor L2 in Accent 2, suggesting a tonal alignment to the vocalic gesture (if /bilder/ remains in the data the significant effect is only present in L2, SE=0.018, t=2.293, p<.05).

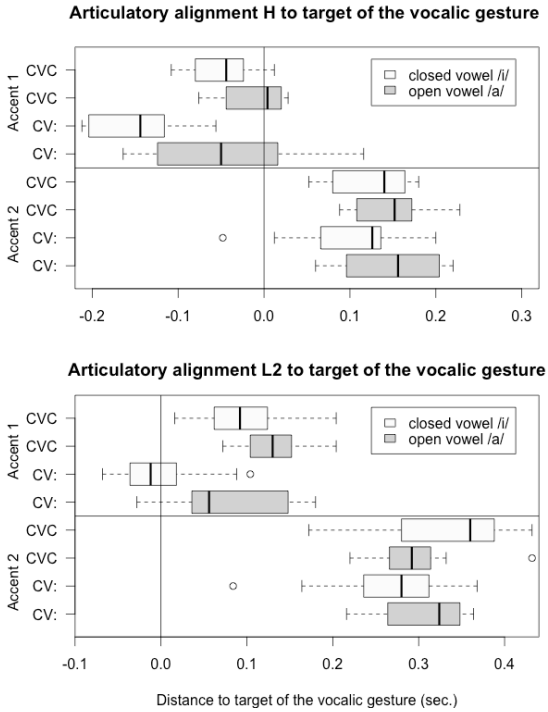

Figure 5. Alignment of H (top) and L2 (bottom) to target of vocalic gesture (vertical line at 0 sec.).

There appears to be no synchronization difference between the two articulatory variables: both articulators seem to differ between the syllable types, but not between

accents (see Figure 6). No significant difference is found between Accent 1 and Accent 2 (SE=0.01, t=-1.193 , *p*=0.235).

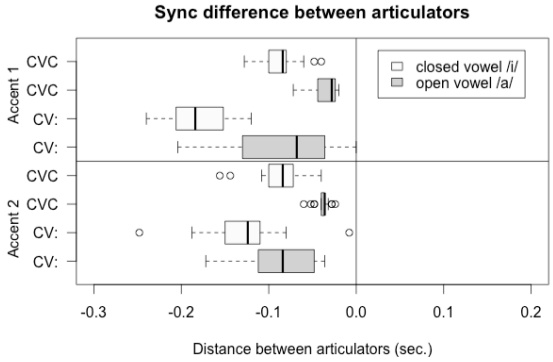

Figure 6. Distance between the articulators. The target of the vocalic gesture is at 0 sec.

Discussion and conclusion

The syllable structure affects the acoustic tonal alignment in both accents but to different units. We interpret this as an alignment of the fall in Accent 1 to the vowel and of Accent 2 to the syllable. This might call for a revision of the Lund Model, which associates both accents with the stressed syllable.

The tonal alignment to the articulatory targets is also affected by syllable type. Moreover, no synchronization difference was found between the accents. These results indicate that the tonal gestures of the accents couple with different articulators: Accent 1 with the consonantal gesture of the lip aperture and Accent 2 to the vocalic gesture of the tongue body. The articulatory, and the acoustic, results indicate that the word accents are different in their phonological nature. It seems plausible that they are separate, which leans more towards the privative than the equipollent distinction.

If the consonantal gesture was to be equivalent to the syllable and the vocalic gesture to the vowel, the articulatory results would contradict the acoustic results, but clearly a more complex relationship is expected, seeing that the relationship between articulatory movements and acoustics is nonlinear. Furthermore, the deviating data on the target word /bilder/ may be explained by coarticulation effects of /i/ and /l/ by speaker SS. It can also be due to the acoustic segmentation, as the deviation is not equally clear in the articulatory alignment.

To be able to further add to the phonological typology and the lexical distinction of the two accents, a follow-up study would benefit from additional segmental structures, as well as measures on the start of the rise. Another aim in future research would be to find a more stable target for the vocalic gesture, but also to include other articulators.

References

Ambrazaitis G & Bruce G (2006). Perception of south Swedish word accents. In: Ambrazaitis G & Schötz S, eds, *Working Papers 52, Proceedings from Fonetik 2006.* Lund, 5-8.

Arvaniti A, Ladd D R & Mennen I (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 26: 3-25.

Atterer M & Ladd D R (2004). On the phonetics and phonology of 'segmental anchoring' of F0: evidence from German. *Journal of Phonetics*, 32: 177-197.

Boersma P & Weenink D (2014). Praat: doing phonetics by computer [Computer program], version 5.4.01. http://www.praat.org/

Browman C P & Goldstein L (1992). Articulatory phonology: an overview. *Phonetica*, 49: 155-180.

Bruce G & Gårding E (1978). A prosodic typology of Swedish dialects. In: Gårding E, Bruce G & Bannert R, eds, *Nordic prosody, Papers from a symposium*. Lund, 219-228.

Bruce G (1998). *Allmän och svensk prosodi*. *Praktisk Lingvistik 16.* Lund: Dept. of linguistics and phonetics.

Bruce G (2007). Components of a prosodic typology of Swedish intonation. In: Riad T & Gussenhoven C, eds, *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*. Berlin, 113-146.

Caspers J & Van Heuven V J (1993). Effects of time pressure on the phonetic realization of the Dutch Accent-lending pitch rise and fall. *Phonetica*, 50: 161-171.

Engstrand O (1997). Phonetic interpretation of the word accent contrast in Swedish: evidence from spontaneous speech. *Phonetica*, 54: 61-75.

Gårding E, Fujimura O, Hirose H & Simada Z (1975). Laryngeal control of Swedish word accents. *Working papers 10*. Lund, 53-82.

Hermes A, Becker J, Mücke D, Baumann S & Grice M (2008). Articulatory gestures and focus marking in German. *Proc. Speech Prosody 2008*, Campinas, 457-460.

Ladd D R, Faulkner D, Faulkner H & Schepman A (1999). Constant 'segmental anchoring' of F0 movements under changes in speech rate. *J. Acoust. Soc. Am*., 106(3): 1543-1554.

Mücke D, Nam H, Hermes A & Goldstein L (2012). Coupling of tone and constriction gestures in pitch accents. In: Hoole P, Bombien L, Pouplier M, Mooshammer C & Kühnert B, eds, *Consonant Clusters and Structural Complexity*. Munich: Mouton de Gruyter, 157-176.

Niemann H, Grice M & Mücke D (2014). Segmental and positional effects in tonal alignment: An

articulatory approach. *Proc. 10th ISSP*, Cologne, 285-288.

Niemann H, Mücke D, Nam H, Goldstein L & Grice M (2011). Tones as gestures: the case of Italian and German. *Proc. ICPhS XVII*, Hong Kong, 1486-1489.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Riad T (2014). *The phonology of Swedish.* Oxford: Oxford University Press.

Riad T (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, 122: 1352-1379.

Riad T (2006). Scandinavian accent typology. In: Viberg Å, ed, *Special issue on Swedish. Sprachtypologie und Universalienforschung (STUF) 59(1).* Berlin, 36-55.

Roll M, Söderström P & Horne M (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, 1520: 116-120.

Svensson Lundmark M (2014). Constant tonal alignment in Swedish word accent II. *Proc. Speech Prosody 7*, Dublin, 987-991.

Xu Y (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica,* 55: 179-203.

An acoustic analysis of the cattle call "kulning", performed outdoors at Säter, Dalarna, Sweden

Anita McAllister[1,2] and Robert Eklund[3]
[1] *Linköping University, Sweden, Division of Speech and Language Pathology, CLINTEC, Karolinska Institute, Stockholm, Sweden*
[2] *Department of Clinical and Experimental Medicine, Division of Speech and Language Pathology, Linköping University, Sweden*
[3] *Department of Culture and Communication, Division of Language and Culture, Linköping University, Sweden*

Abstract
This paper summarizes recent research on 'kulning', a surprisingly understudied Swedish cattle call singing style. In a previous study (Eklund, McAllister & Pehrson, 2013), we compared kulning and head voice ('falsetto') as recorded in a normal room and in an anechoic chamber. This paper reports from an analysis of the same "kulning" song recorded outdoors on location in Säter, Dalarna (Sweden), close to the singer's home, which makes the data more ecologically valid and allows comparisons between "clean" indoor recordings and more authentic outdoor recordings. Several recordings were made, but the present article analyses recordings made simultaneously at 1 meter and 11 meters from the singer. Results indicate that for the vowels [a] and [ɥ̟] partials in kulning, as compared to head voice, are visible at both higher frequencies and at a longer distance, which provides an acoustic rationale for the development of the singing style, intended to be heard at a long distance.

Introduction

The Swedish cattle call singing style 'kulning' is surprisingly understudied, despite its almost mythical status in Swedish folklore. Throughout history, long-distance calls have been created at several different locations where there has been a need of making oneself heard over long distances. Kulning is the most common term for a specific type of cattle or herding calls used mainly in the provinces Dalarna, Härjedalen and Jämtland (all in Sweden) and is used to call cows or goats when it is time to be milked. In Eklund, McAllister & Pehrson (2013) we compared kulning and head voice recorded in a normal room and in an anechoic chamber.

The present paper summarizes results in two recent publications (Eklund & McAllister, 2015a, 20015b) where the vowels [*a*] and [ɥ̟] were recorded in an outdoors setting in Säter (Dalarna, Sweden), close to the singer's home, yielding more ecologically valid data. The data consisted of simultaneous recordings at two distances from the singer, 1 and 11 meters.

For an account of previous research, the reader is referred to Eklund and McAllister (2015a, 2015b).

Data collection and method

The singer (FP) – the same singer as in Eklund, McAllister & Pehrson, 2013 – is educated in kulning at Musikkonservatoriet in Falun and Malungs Folkhögskola, and by Agneta Stolpe and Ann-Sofi Nilsson. Data consisted of FP singing the same cattle call as in our previous study and was recorded in both kulning and head voice (sometimes incorrectly referred to as "falsetto") modes. The present study is based on simultaneous recordings made at 1 and 11 meters from the singer, using the same kind of equipment. All recordings were made on 7 September 2013, on location in Säter, Dalarna in Sweden, close to where the singer grew up. The microphones used in this study were two Shure Pro Beta 58A that independently fed into two high-definition Canon HG-10 video cameras. Air humidity at the time of recording was around 70%, the temperature was around 21 degrees Celsius while wind speed was around 10 km/h (data from http://freemeteo.com). All data were resampled to 44.1 kHz, 16 bit, mono, using TMPGEnc 4.0 XPress. Acoustic analyses were carried out using Cool Edit Pro 2.0, Cool Edit 2000, WaveSurfer 1.8.8p4 and Praat 5.3.84.

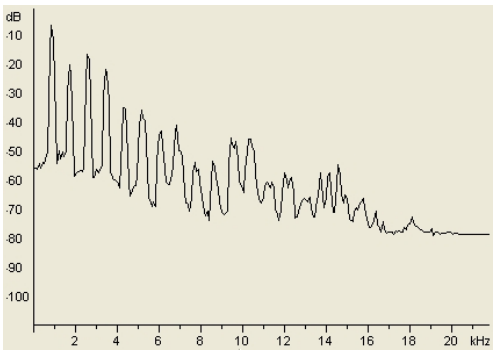

Figure 1a: Kulning [a] at 1 meter. LTAS/FFT/Hamming analysis.

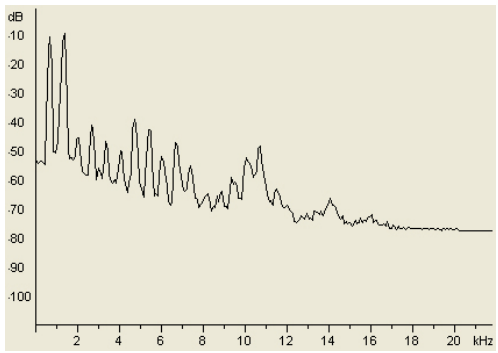

Figure 1b: Head voice [a] at 1 meter. LTAS/FFT/Hamming analysis.

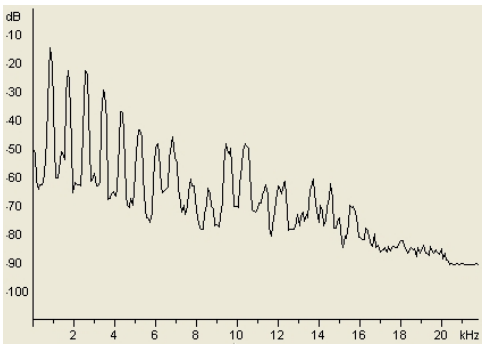

Figure 2a: Kulning [a] at 11 meters. LTAS/FFT/Hamming analysis.

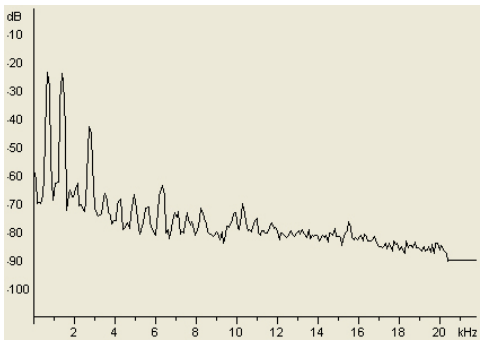

Figure 2b: Head voice [a] at 11 meters. LTAS/FFT/Hamming analysis.

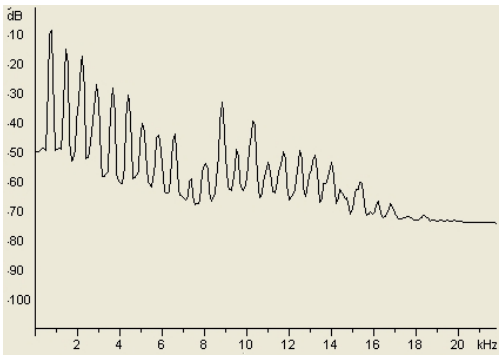

Figure 3a: Kulning [ʉ̬] at 1 meter. LTAS/FFT/Hamming analysis.

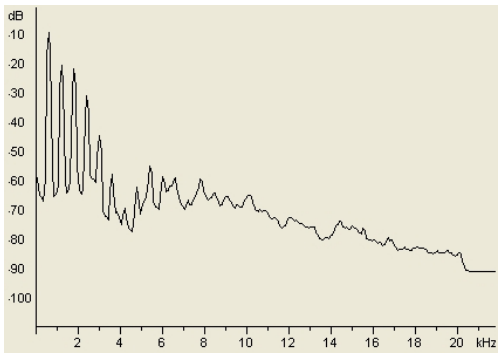

Figure 3b: Head voice [ʉ̬] at 1 meter. LTAS/FFT/Hamming analysis.

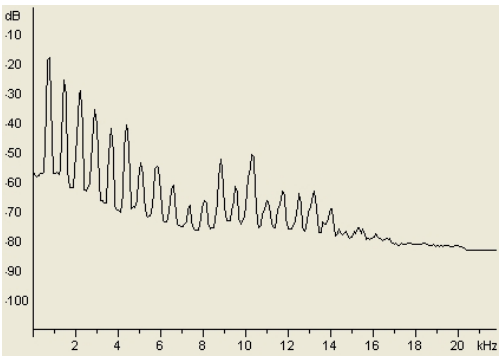

Figure 4a: Kulning [ʉ̬] at 11 meters. LTAS/FFT/Hamming analysis.

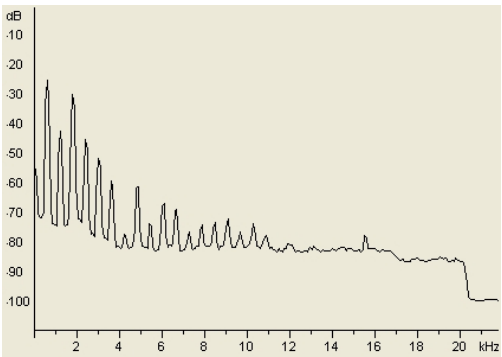

Figure 4b: Head voice [ʉ̬] at 11 meters. LTAS/FFT/Hamming analysis.

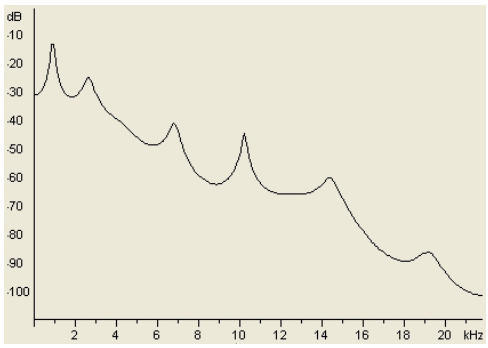

Figure 5a: Kulning [a] at 1 meter. LTAS/LPC/Hamming analysis.

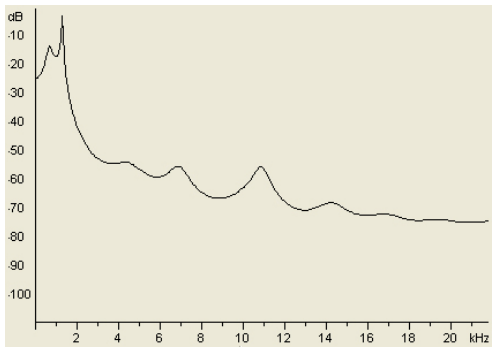

Figure 5b: Head voice [a] at 1 meter. LTAS/LPC/Hamming analysis.

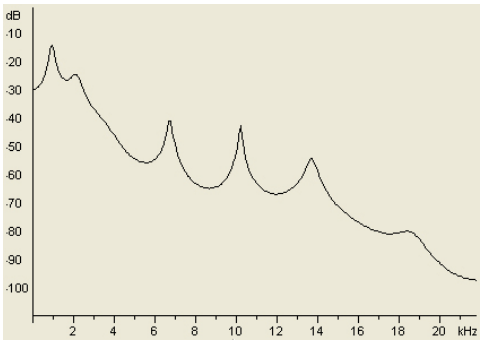

Figure 6a: Kulning [a] at 11 meters. LTAS/LPC/Hamming analysis.

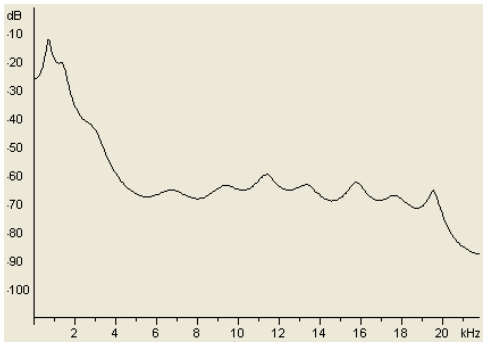

Figure 6b: Head voice [a] at 11 meters. LTAS/LPC/Hamming analysis.

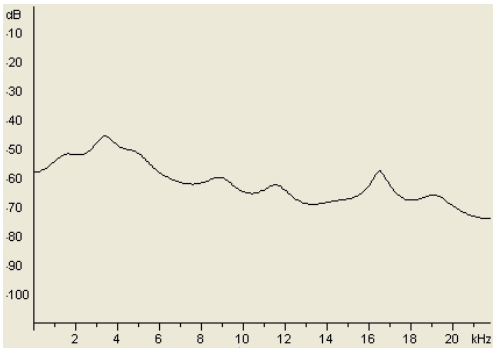

Figure 7a: Kulning [ʉ̬] at 1 meter. LTAS/LPC/Hamming analysis.

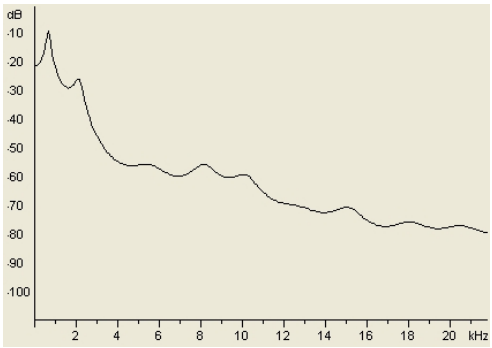

Figure 7b: Head voice [ʉ̬] at 1 meter. LTAS/LPC/Hamming analysis.

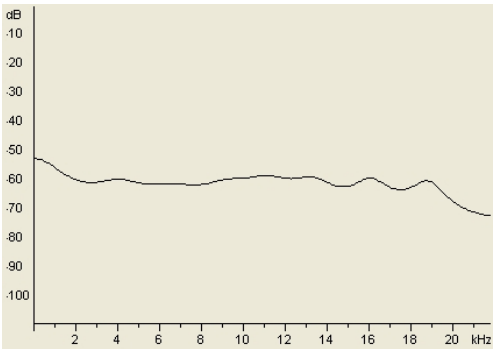

Figure 8a: Kulning [ʉ̬] at 11 meters. LTAS/LPC/Hamming analysis.

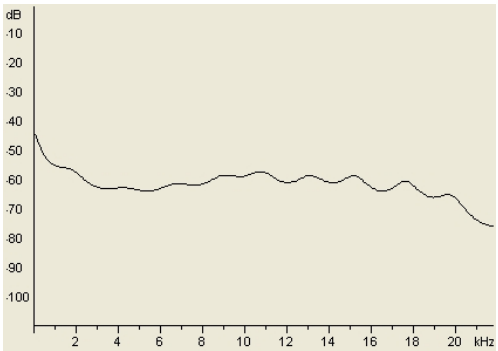

Figure 8b: Head voice [ʉ̬] at 11 meters. LTAS/LPC/Hamming analysis.

Results

Analysis material

In order to match the data in our previous study, we analysed the vowels [*a*] and [ʉ̟]. Both vowels were excised from the recordings and Long-Term Average Spectra were created for both partials (FFT) and formant (LPC) analyses.

LTAS/FFT/partials

The results from the FFT analyses are shown in Figures 1a through 4b.

SPL levels at one meter were 84.2 dB for kulning and 81.5 dB for head register and at 11 meters 74.8 and 56.3 respectively. The comparative lack of attenuation in kulning is striking. In general more partials are seen in the kulning rendition where partials are clearly visible up to 16 kHz for both vowels. However, this difference is not as striking as in our indoor recordings, where the head voice version only had visible partials up to around 6 kHz. Here, partials can be observed up to at least 11 kHz in head voice for both vowels.

LTAS/LPC/formants

The results from the LPC analyses are shown in Figures 5a through 8b.

The most striking observation is that for the vowel [*a*], curves are very similar regardless of distance in kulning. Formants are more or less unperturbed by adding an additional ten meters distance from the source. For the vowel [ʉ̟] results are somewhat different since no clear formant peaks can be detected in the kulning style. However, in head voice, the first formant is clearly visible. The difference between the two singing styles is most apparent at one meter from the singer; see Figures 7a and 7b.

Discussion and conclusions

In this paper, we report on the analyses of two vowels obtained from outdoor kulning singing. By and large, our previous observations were replicated, showing that kulning, compared to head voice, attenuates far less with distance than head voice. Kulning also has clearer partials at higher registers, and preserves partials patterns almost unperturbed at the greater distance of 11 meters, as compared to 1 meter, from the source. Partials were clearly observed for both vowels up to ~16 kHz at both distances.

Regarding formant patterns, those were preserved at 11 m for the vowel [*a*] but to a lesser degree for the vowel [ʉ̟] where flat LTAS spectra were found both at one and 11 meters. Of interest is the very flat formant curve, compared to *[a]*, where formants are still discernible. However, it should be noted that it is difficult both to produce and perceive "exact" vowel qualities when singing at high frequencies, which means that our chosen IPA characters can be considered somewhat arbitrary.

The main difference in partials and formant slopes between the two singing modes is the preserved loudness in kulning. This does in turn help explain why the singing mode was developed to call cattle that could be at considerable distances from the singer.

Future studies

Planned future studies include inverse filter analyses, fiberoptic endoscopic and electro-glottographic examinations in new data collections. We hope that such studies will help elucidate the underlying glottal configuration that produces the observed distance-resilient formants, with visible partials up to ~16 kHz and less loudness attenuation with distance, compared to the somewhat similar-sounding head voice alternative.

Acknowledgments

The authors would like to thank our singer Fanny Pehrson, who happily volunteered to be the subject of our studies. Thanks also to Ahmed Geneid for interesting comments.

References

Eklund, R & McAllister, A (2015a). An acoustic analysis of 'kulning' (cattle calls) recorded in an outdoor setting on location in Dalarna (Sweden). *Proceedings of ICPhS 2015*, 10-14 August 2015, Glasgow, Scotland.

Eklund, R & McAllister, A (2015b). An acoustic analysis of 'kulning' (cattle calls) recorded in an outdoor setting on location in Dalarna (Sweden). Part II: the case of [a]. Submitted for publication.

Eklund, R, McAllister, A & Pehrson, F (2013). An acoustic comparison of voice characteristics in 'kulning', head and modal registers. In: Robert Eklund (ed.), *Proceedings of Fonetik 2013, the XXVIth Swedish Phonetics Conference, Studies in Language and Culture*, no. 21, 12–13 June 2013, Linköping University, Linköping, Sweden.

Agonistic vocalisations in domestic cats: a case study

Susanne Schötz
Dept. of Logopedics, Phoniatrics and Audiology, Lund University

Abstract

Introducing a new cat to a home with resident cats may lead to stress, aggression and even fights. In this case study 468 agonistic cat vocalisations were recorded as one cat was introduced to three resident cats in her new home. Six vocalisation types were identified: growl, howl, howl-growl, hiss, spit and snarl. Numerous other intermediate and complex vocalisations were also observed. An acoustic analysis showed differences within and between all types. Future studies include further acoustic analyses of cat vocalisations produced by a larger number of cats.

Introduction

The cat (*Felis catus,* Linneaus 1758) was domesticated about 10,000 years ago, and is now one of the most popular pets of the world with more than 600 million individuals (Driscoll et al., 2009; Turner & Bateson, 2000). Domestic cats have developed a more extensive, variable and complex vocal repertoire than most other members of the Carnivora, which can be explained by their social organisation, their nocturnal activity and the long period of association between mother and young (Bradshaw, 1992). Still, we know surprisingly little about the phonetic characteristics of these sounds. The few existing studies of cat vocalisations report findings from only a limited number of cats, vocalisation types, or methods (e.g. Moelk, 1944; Brown et al., 1978; McKinley, 1982; Shipley et al., 1988, 1991; Farley et al., 1992, Nicastro & Owren 2003, Yeon et al., 2011).

Cat vocalisations are generally divided into three categories: sounds produced with the mouth closed, sounds produced with an opening–closing mouth, and sounds produced with the mouth held tensely open (Moelk, 1944; McKinley, 1982). Many previous studies have focused on purring, on human– or prey–directed cat sounds or on kitten vocalisations (e.g. Moelk, 1944; Brown et al., 1978; Nicastro & Owren 2003; Eklund et al., 2010). There are few studies on agonistic sounds. Yeon et al. (2011) found that non-socialised cats produced far more aggressive and defensive calls than socialised cats. Brown et al. (1978) studied mother-kitten interaction and found that kittens produced howls and growls at about 3-4 weeks of age, and hissing and spitting around the age of 30 days.

The main purpose of this study was to learn more about the acoustic-phonetic characteristics of agonistic cat vocalisations. In earlier studies, I have recorded and analysed only non-agonistic vocalisations of my own cats (Schötz & Eklund, 2011, Schötz, 2012, 2013, Schötz & van de Weijer, 2014). However, an opportunity to record agonistic sounds came when I introduced my new cat to my three previous cats. In this small case study I wanted to find out 1) what agonistic sounds the cats produced, 2) what types were the most frequent ones, and 3) what their acoustic-phonetic features in terms of duration, F_0 and spectral centroid were.

Agonistic cat vocalisations

Vocal communication between cats is largely restricted to three types of interactions; mother-young, sexual and agonistic. Agonistic sounds are aggressive and defensive sounds used to warn, shock or startle an intruder or attacker. Most agonistic sounds are strained-intensity calls, produced while the cat is tensing its whole body in preparation for a fight. They are often used in combination with visual body posture signals, both attempting to persuade an opponent that the cat is bigger than it really is. For instance, cats can combine a low pitched growl or a long yowl with drawing themselves up to their full height, turning partially sideways and making their hair stand on end (Bradshaw & Cameron-beaumont, 2000). Several types of agonistic sounds have been described in earlier literature, including the growl, the howl, the snarl, the hiss, and the spit.

Growl

The growl is a guttural, harsh, regularly and rapidly pulse-modulated, low-pitched (100-225 Hz) sound of usually long duration. It is produced during a slow steady exhalation while the mouth is held slightly open in the same position (Moelk, 1944; McKinley, 1982; Bradshaw & Cameron-beaumont, 2000; Beaver, 2003; Eklund et al., 2012, Bradshaw, 2013). Brown et al. (1978:556) describe growls as largely fricative and long in duration. The growl is transcribed as [grrr..] with a vocalic [rrr...] or rhotic [ʌ], occasionally beginning with an [m] by Moelk (1944) and sounds a bit like a prolonged low pitched English alveolar approximant or retroflex produced with creaky voice; [ɻ̰]. It is mainly used to signal danger or to warn or scare off an opponent. Variations in duration and F_0 are common, and often the growl is either intertwined with howls/moans/yowls and hisses, or an intermediate vocalisation between e.g. a growl and a howl. Growls during a fight may vary between 400 and 800 Hz in F_0 (Houpt, 2004).

Howl, moan, yowl, anger wail

Howls, moans, yowls, or anger wails are long and often repeated vocalic warning signals usually produced by gradually opening the mouth wider and closing it again. During a threatening situation, they are often merged or combined with by growls in long sequences with slowly varying F_0 and intensity (Brown et al., 1978; Eklund et al., 2012). Moelk (1944) transcribes the anger wails [wɑːouː], with the vowel intensifying toward [æ], and points out that "[s]lighter wailing [...] occurs occasionally in connection with the growl in highly annoying situations which do not lead to fighting". Brown et al. (1978:566) found howls to be tonal sounds occurring in threatening or defensive responses with a wide variation in frequency distribution and modulation. Moans are described by McKinley (1982) as long, often slowly frequency-modulated vowel sounds of "o" or "u" occurring in the same situations as the growls. Bradshaw & Cameron-beaumont (2000) distinguish howls from yowls in that howls are typically shorter in duration (howls: 0.8-1.5 s., yowls 3–10 s.), and higher in F_0 (howls: 700 Hz, yowls 200–600 Hz).

Snarl and pain shriek

Snarls and pain shrieks are loud, harsh and high-pitched vocalisations produced during active fighting (McKinley, 1982:13). Snarls are used to startle or scare an opponent, and are described by Moelk (1944) as "rapid inhalations harshly vocalised and marked by a heavy initial intake of breath and stopped suddenly with a slight [o] sound, [ˈœːo]". Pain shrieks are short intense cries of tense vowels, often [æ], [ɛ] or [i], and are characterised by "great strain at mouth and throat and the force of breath" (Moelk, 1944).

Hiss and spit

Hissing and its more intense variation spitting are involuntary reactions to when a cat is surprised by an (apparent) enemy. The cat changes position with a startle and breath is being forced rapidly through the slightly open mouth before stopping suddenly; [fft!] (Moelk, 1944:194). McKinley (1982) describes the hiss as an "agonistic vocalization given with the mouth wide open teeth exposed, and sounding like a long exhalation", and the spit as "a very short explosive sound, given in agonistic situations frequently before or after a hiss".

Other sounds

Occasionally other sounds are produced in (apparent) agonistic situations. These include chirps, meows and chirrups. Cats usually chirp at birds or insects (prey), and dominant cats may chirp at the sight of an inferior or smaller cat. Meows can be produced during play with other cats, or if a situation is perceived as playful rather that threatening by one of the cats. Moreover, distinctive coaxing calls or chirrups may be used by tomcats to lure young or neutered males out of their homes to fight.

Method

Preparation and data collection

Vimsan (V, female, about 2 years old) was found outside our home injured in October 2014, and after recovering she was slowly introduced to the other cats of her new home. The first few weeks she was confined to an area of the house to which the other three cats Donna, Rocky and Turbo (D, R and T; 1 female, 2 males, all 4.5 year old siblings from the same litter) had no access. They were, however, slowly given increased opportunities to smell blankets and

toys that had been used by the other cat(s), and then allowed into each others areas without the resident cat(s) being present, and after a week they were able to look at and smell each other through a narrow opening of a door. When V was let out to the other cats for the first time on November 13, 2014, I began recording the cats' vocalisations several hours every day and I continued for eight days until the cats had become so used to one another that they hardly used any agonistic vocalisations anymore.

The equipment consisted of a Sony HDR-CX730E video camera recorder with a Sony ECM-CG50 electret condenser shotgun microphone. Additional recordings were done with an Apple iPhone 3G. All recordings were transferred to a computer (Wave, 44,1 kHz/16 bit) for further analysis.

Table 1. The six agonistic vocalisation types recorded in the study.

Type	*Descriptive term*
Gr	*Growl*
Ho	*Howl, moan, yowl, anger wail*
HoGr	*Combination of howl(s) and growl(s)*
Hs	*Hiss*
Sn	*Snarl*
Sp	*Spit*

Table 2. Number of vocalisations of the four cats (D, R, T, V) by vocalisation type (for type descriptions, see Table 1).

Cat	*Gr*	*Ho*	*HoGr*	*Hs*	*Sn*	*Sp*	*Total*
D	13	175	114	38	3	22	*365*
R	2	47	1	4	0	2	*56*
T	13	2	4	7	0	1	*27*
V	3	6	0	5	4	2	*20*
Total	*31*	*230*	*119*	*54*	*7*	*27*	*468*

Preprocessing, categorisation and analysis

All recordings were transferred to a computer, and audio files (wav, 44.1 kHz, 16 bit, mono) extracted. The waveforms were normalised for amplitude and the vocalisations segmented and labelled in *Praat* (Boersma & Weenink, 2014). Out of 516 recorded vocalisations 48 were discarded as they were non-agonistic (chirps, meows), too weak in intensity or contained overlapping sounds. The remaining 468 agonistic vocalisations were categorised into six types (see Table 1) and used in the acoustic analysis. Measures of duration as well as of F_0 for the voiced sounds, and of centre of gravity (centroid or spectral mean) for the voiceless sounds were obtained with a *Praat* script and manually checked. The acoustic results were then further analysed and summarised using R (R Core Team, 2015). The six vocalisation types are listed in Table 1. Figure 1 and Table 2 display the number and proportion of vocalisations of each type by the four cats.

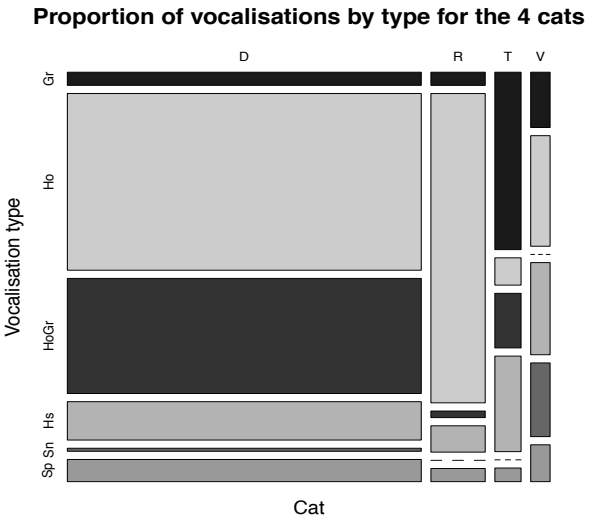

Figure 1. Mosaic plot of the proportions of the six vocalisation types growl *(Gr),* howl *(Ho),* howl-growl *(HoGr),* Hiss *(Hs),* Snarl *(Sn) and* Spit *(Sp) for the four cats (D, R, T, V).*

Results

D was by far the most vocal cat of this case study with a total of 365 vocalisations. R produced only 56, T 27 and V 20 sounds. Not all cats produced all six types of vocalisations. The most frequent vocalisation type was howl with 230 tokens, followed by howl-growl (119 tokens), hiss (54 tokens), growl (31 tokens), spit (27 tokens) and snarl (7 tokens). The results of the acoustic analysis of the six agonistic vocalisation types are described below. Median values were very close to mean values, and therefore only mean values are presented here.

Growl

The growls were often low [ɪ̰]-like sounds with fairly level F_0, around 70-200 Hz. Growls were produced as warnings as one cat came too close to one of the other cats. Some tokens seemed to be produced with falsetto voice quality with higher F_0. Durations varied between 0.83 and 4.46 sec, with an overall mean of 2.50 sec. These values, as well as individual values for each cat, are shown in Table 3. Figure 2 shows the waveform, broadband spectrogram and F_0 contour of a typical growl.

Table 3. Mean durations, as well as minimum, maximum and mean F_0 of growl (Gr).

Cat	*meanDur*	*minF_0*	*maxF_0*	*meanF_0*
D	2.27 s	128 Hz	475 Hz	285 Hz
R	1.15 s	70 Hz	78 Hz	73 Hz
T	2.77 s	46 Hz	482 Hz	283 Hz
V	3.25 s	70 Hz	99 Hz	79 Hz
All	*2.50 s*	*46 Hz*	*482 Hz*	*250 Hz*

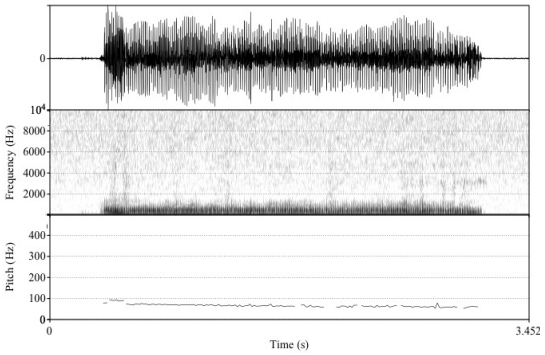

Figure 2. Example waveform, broadband (300 Hz) spectrogram and F_0 contour of growl (Gr).

Howl, moan, yowl, anger wail

The howls and similar sounds recorded in this case study varied in duration between 0.22 and 8.79 sec. They often displayed a tonal rising-falling pattern with an F_0 ranging from 128 to 842 Hz, and also varied in their vowel quality. Closed vowel qualities like [ɪ], [ɨ] or [ʏ] as well as diphthongs like [au], [ɛɔ] or [ɑo] were common, but also semivowel qualities were observed. Howls were uttered as warnings and often accompanied by growls and howl-growls in long sequences of repetitions as one cat had moved too close to an opponent. Figure 3 displays the waveform, broadband spectrogram and F_0 contour of an example howl, and numeric values for this type are shown in Table 4.

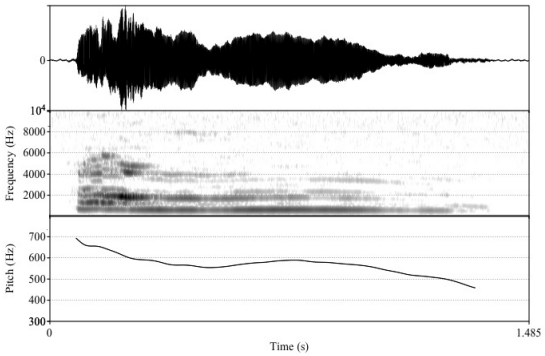

Figure 3. Example waveform, broadband (300 Hz) spectrogram and F_0 contour of howl (Ho).

Table 4. Mean durations, as well as minimum, maximum and mean F_0 of Howl (Ho).

Cat	*meanDur*	*minF_0*	*maxF_0*	*meanF_0*
D	2.27 s	128 Hz	692 Hz	285 Hz
R	1.82 s	241 Hz	797 Hz	624 Hz
T	0.87 s	263 Hz	579 Hz	367 Hz
V	1.64 s	603 Hz	842 Hz	769 Hz
All	*1.90 s*	*143 Hz*	*907 Hz*	*684 Hz*

Howl-growl combinations and transitions

Combinations of howls and growls were produced mainly by D in this case study, although R and T uttered a few tokens of this type. These sounds were used in similar contexts as growls and howls, and were between 1.30 and 9.47 sec. in duration. F_0 typically increased and decreased with the transitions from howl to growl and ranged between 56 and 974 Hz. Figure 4 shows a typical howl-growl example, and Table 5 displays the numeric values for the acoustic analysis of this vocalisation type.

Table 5. Mean durations, as well as minimum, maximum and mean F_0 of howl-growl (HoGr).

Cat	*meanDur*	*minF_0*	*maxF_0*	*meanF_0*
D	3.66 s	121 Hz	974 Hz	519 Hz
R	1.93 s	151 Hz	459 Hz	274 Hz
T	6.28 s	56 Hz	837 Hz	324 Hz
V	-	-	-	-
All	*3.73 s*	*56 Hz*	*974 Hz*	*510 Hz*

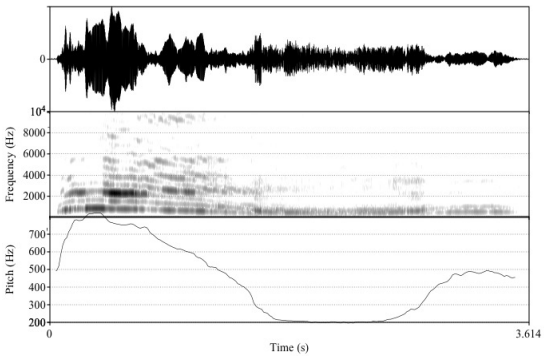

Figure 4. Example waveform, broadband (300 Hz) spectrogram and F_0 contour of howl-growl (HoGr).

Snarl and pain shriek

Snarls and pain shrieks were produced only during actual fights, and were harsh, short and loud calls with durations ranging from 0.19 to 0.64 sec, and F_0 varying between 301 and 521 Hz, as shown in Table 6. Vowel qualities included [a] and [æ]. As these sounds occurred only in actual fights between the two female cats, it was impossible to judge which of the cats

produced the vocalisations, and although my guess is that D produced three and V four snarls, it is possible that V produced all of them, as they are very similar in voice quality and F_0. Figure 5 shows an example of a snarl, and numeric values for this type are shown in Table 6.

Table 6. Mean durations, as well as minimum, maximum and mean F_0 of snarl (Sn).

Cat	*meanDur*	*minF_0*	*maxF_0*	*meanF_0*
D	0.46 s	301 Hz	521 Hz	461 Hz
R	-	-	-	-
T	-	-	-	-
V	0.34 s	301 Hz	521 Hz	464 Hz
All	*3.73 s*	*56 Hz*	*974 Hz*	*510 Hz*

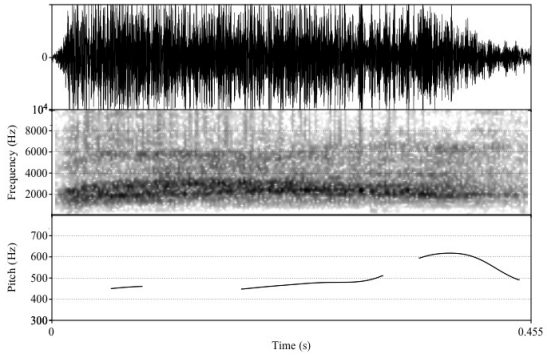

Figure 5. Example waveform, broadband (300 Hz) spectrogram and F_0 contour of snarl (Sn).

Hiss and spit

Hisses and spits are voiceless vocalisations. The hisses produced by the cats in this case study were uttered as warnings and sounded a bit like the fricatives [ʂ], [ç], or [h]. Spits (hisses) were used as intense warnings or to shock an opponent and often sounded more like affricates [tʂ] or [tç]. However, hisses and spits were not always easy to tell apart, as they sometimes sounded very similar. Hisses (0.42–1.05 sec.) were generally longer than spits (0.27–0.70 sec.), with a lower centre of gravity, as shown in Table 7. Centre of gravity standard deviations varied between 2080 and 2507 Hz, suggesting a wide dispersion of the noise energy in both types. Figure 6 shows the waveform and spectrogram of one hiss and one spit.

Table 7. Mean durations (mDur) and centres of gravity (cog) of Hiss (Hs) and Spit (Sp).

Cat	*mDurHs*	*cogHs*	*mDurSp*	*cogSp*
D	0.68 s	1186 Hz	0.52 s	2116 Hz
R	0.69 s	820 Hz	0.62 s	1506 Hz
T	0.80 s	937 Hz	0.55 s	1464 Hz
V	0.66 s	957 Hz	0.62 s	1562 Hz
All	*0.70 s*	*1105 Hz*	*0.54 s*	*2006 Hz*

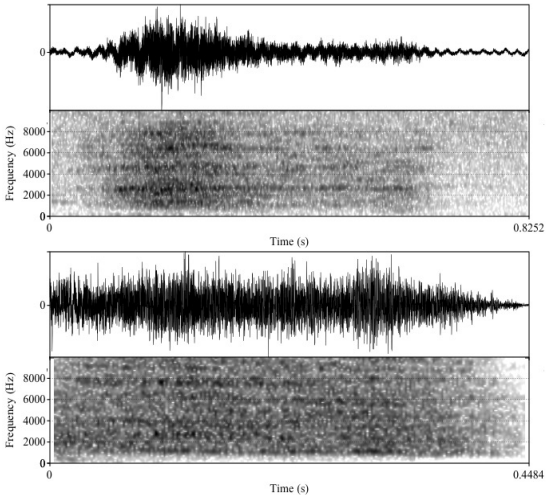

Figure 6. Example waveforms and broadband (300 Hz) spectrogram of hiss (HS) (top) and Spit (sp) (bottom).

Discussion and future work

In this case study, the main purpose was to find out what types of agonistic vocalisations were used by the participating cats, and what some of their acoustic-phonetic features were. From the 468 analysed tokens, at least six different vocalisation types were identified. Moreover, several intermediate patterns between e.g. hiss and spit, snarl and pain shriek, and between howl and growl were not uncommon. Such sounds were harder to subdivide into types. Furthermore, complex vocalisations, including combinations of howls and growls, growls and spits were observed. In futures studies, intermediate and complex vocalisations will be analysed in greater detail.

Most of the vocalisations (78%) were produced by one cat (D), who was the most active and aggressive cat of this case study. Still, the fact that so many agonistic types were identified suggests that cats are able to vary their vocal signals to a large extent even in such a narrowly defined behavioural context. This is in line with Moelk (1944:185), who found that the vocal repertoire of the domestic cat is characterised by "an indefinitely wide variation of sound and of patterning".

Large variations in F_0 and durations within and between the different types were also found. Especially howls and howl-growls comprised a large number of intonation patterns, which is in line with previous studies (Schötz, 2012; 2013; 2014). It is possible that cats use variations in F_0 to signal paralinguistic information. This will be studied further in future experiments.

Agonistic vocalisations are not easy to record naturally without human–cat interaction. The present study used a particular case where one new cat was introduced to three cats already living in her new home and without any human–cat interaction. This method was found to be adequate for recording cat–cat agonistic vocalisations, and will be used again if possible in future studies with other cats. In this study none of the cats were forced to do anything against their will. They could retreat to a safe place whenever they wanted to (and they often did). After two weeks of mainly aggressive and defensive behaviour, the cats calmed down, and they now seem to tolerate each other and are getting along better.

The results of this pilot study should be regarded as tentative, due to the often limited number of tokens analysed of each type. Future work includes a larger study of cat vocalisations, including intonation and an initial formant analysis of the different vocalisation types, especially the vowels.

Acknowledgements

A warm thanks goes to Donna, Rocky, Turbo and Vimsan for their patient participation in this case study. [m̩hrn]!

References

Beaver B V (2003). *Feline Behaviour: A Guide for Veterinarians*. W. B. Saunders: ST. Louis, MO.

Boersma, P, Weenink, D (2014). *Doing phonetics by computer* [Computer program]. Version 5.4.01, retrieved 9 November 2014 from http://www.praat.org/.

Bradshaw, J (1992). *The Behaviour of the Domestic Cat*. Redwood Press, Bristol.

Bradshaw, J, Cameron-beaumont, C, (2000). The signalling repertoire of the domestic cat and its undomesticated relatives. In: Turner, D C, Bateson, P (Eds.), *The Domestic Cat: The Biology of its Behaviour*. Cambridge University Press, Cambridge, 67–793.

Bradshaw, J (2013). *Cat Sense: The Feline Enigma Revealed*, London: Allen Lane.

Brown, K A, Buchwald, J S, Johnson, J R, Mikolich, D J (1978). Vocalization in the cat and kitten. *Developmental Psychobiology*, 11: 559–570.

Driscoll, C A, Clutton-Brock, J, Kitchen, A C, O'Brien, S J (2009). The taming of the cat. *Scientific American,* June 2009, 68–75.

Eklund, R, Peters, G, Duthie, E D (2010). An acoustic analysis of purring in the cheetah (*Acinonyx jubatus*) and in the domestic cat (*Felis catus*). In: *Proceedings of Fonetik 2010*, Lund University, 2–4 June 2010, Lund, 17–22.

Eklund, R, Peters, G, Weise, F, Munro, S (2012). A comparative acoustic analysis of purring in four cheetahs. In: *Proceedings from FONETIK 2012*. Gothenburg, 2012, 37–40.

Farley, G R, Barlow, S M, Netsell, R, Chmelka, J V, (1992). Vocalizations in the cat: behavioral methodology and spectrographic analysis. *Exp. Brain Res*. 89: 333–340.

Houpt, K (2004). *Domestic Animal Behavior for Veterinarians and Animal Scientists*, 4th edition. Blackwell Publishing: Ames, IA.

McKinley, P E (1982). *Cluster analysis of the domestic cat's vocal repertoire*. Unpublished doctorial dissertation. University of Maryland, College Park.

Moelk, M (1944) Vocalizing in the House-Cat; A Phonetic and Functional Study. *The American Journal of Psychology*. 57:2: 184–205.

Nicastro, N, & Owren, M J (2003). Classification of domestic cat (Felis catus) vocalizations by naïve and experienced human listeners. *Journal of Comparative Psychology*, 117: 44–52.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Shipley, C, Buchwald, J S, Carterette, E C. (1988). The role of auditory feedback in the vocalization of cats. *Exp. Brain Res*. 69: 431–438.

Shipley, C, Carterette, E C, Buchwald, J S (1991). The effects of articulation on the acoustical structure of feline vocalizations. *J. Acoust. Soc. Am*. 89: 902–909.

Schötz, S, Eklund, R (2011). A comparative acoustic analysis of purring in four cats. In *Proceedings of Fonetik 2011*, Speech, Music and Hearing, KTH, Stockholm, TMH-QPSR, 51: 9–12.

Schötz, S (2012). A phonetic pilot study of vocalisations in three cats. In *Proceedings of Fonetik 2012*, Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, 45–48.

Schötz, S (2013). A phonetic pilot study of chirp, chatter, tweet and tweedle in three domestic cats. In *Proceedings of Fonetik 2013*, Linköping University, 65–68.

Schötz, S & van de Weijer, J (2014). A Study of Human Perception of Intonation in Domestic Cat Meows. In *Proceedings of Speech Prosody 2014*, Dublin.

Schötz, S (2014). A pilot study of human perception of emotions from domestic cat vocalisations. In *Proceedings of Fonetik 201*, Department of Linguistics, Stockholm University, 95–100.

Turner, D C & Bateson, P eds. (2000). *The domestic cat: the biology of its behaviour*, 2nd edn. Cambridge University Press, Cambridge.

Yeon, S C, Kim, Y K, Park, S J, Lee, S S, Lee, S Y, Suh, E H, Houpt, K A, Chang, H H, Lee, H C, Yang, B G, Lee, H J (2011). Differences between vocalization evoked by social stimuli in feral cats and house cats, *Behavioural Processes* 87, Issue 2, 183–189.

Using tonal cues to predict inflections

Pelle Söderström, Merle Horne and Mikael Roll
Department of Linguistics and Phonetics, Lund University

Abstract

The present article discusses the Central Swedish word accents – Accent 1 and Accent 2 – and their productive association to suffixes from the point of view of prediction theories of speech processing. Based on recent neurophysiological findings, we propose that both word accents are used predictively, but that Accent 1 is more 'predictively useful' than Accent 2, due to the fact that Accent 1 stems signal a smaller well-defined set of upcoming affixes as compared to Accent 2. This 'usefulness' allows suffixes to be pre-activated before they have even been heard.

Introduction

The brain constantly makes predictions about upcoming events, across many levels of sensory processing (Bar, 2007; Friston, 2010). These predictions allow us to streamline cognitive processing and increase its efficiency (van Boxtel & Böcker, 2004; Skipper, 2015). The present article reviews research into predictive mechanisms in language processing, focusing on the Central Swedish word accents: Accent 1 (a low tone on the stressed word stem vowel) and Accent 2 (a high tone on the stem vowel). The word accents are associated with suffixes in the mental lexicon (Rischel, 1963; Bruce, 1977; Riad, 2012), so that e.g. a noun like *bil* 'car' is associated with Accent 1 if the singular suffix *-en* is attached to the word stem, or Accent 2 if the plural suffix *-ar* is attached (*bil*$_1$*-en*/*bil*$_2$*-ar*). Furthermore, Accent 2 is associated with compound words in Central Swedish (such as *bildäck*$_2$ 'car tyre'), leading to the assumption that Accent 2 stems can activate more word forms as compared to Accent 1 stems.

The highly productive association between word accents and endings makes Swedish an excellent candidate for studying rapid, online predictions about upcoming language structures. While both word accents seem to be used predictively, there are differences in the way they are used in speech processing and these differences are currently the target of further investigation. Specifically, Accent 1 has been claimed to be more "useful" for prediction than Accent 2 (Roll, Söderström, Mannfolk, Shtyrov, Johansson, van Westen & Horne, submitted), due to its being associated with fewer word forms. The present article will present findings from a recent study which supports that claim and which sheds further light on the processing differences between Accent 1 and 2.

Theoretical assumptions and previous findings

Using behavioural and neurophysiological methods, several studies have investigated the online processing of word accents (Roll, Horne & Lindgren, 2010; Söderström, Roll & Horne, 2012; Roll, Söderström & Horne, 2013; Roll et al., submitted; Söderström, Horne & Roll, submitted). All of these studies have reached the conclusion that the productive association between word accents and morphology is related to speakers' ability to predict an upcoming suffix based on the word accent. In event-related potential (ERP) investigations, mismatching combinations of word accents and suffixes elicit longer response times and reprocessing effects at suffix onset. One such effect is the 'P600', a positive-going ERP deflection found for various types of violations in cognitive processing. In addition to being associated with e.g. syntactic violations in language processing, it has also been proposed that the P600 reflects more general processes related to violated/disconfirmed predictions (e.g. van de Meerendonk, Kolk, Vissers & Chwilla, 2008). Thus, finding a P600 effect for mismatching tone-suffix combinations could be suggestive of a rapid prediction (from the stem to the onset of the suffix) that has been made and disconfirmed.

However, on its own, the P600 is a relatively late effect which could also be claimed to not have any direct bearing on earlier potentially predictive mechanisms. It is possible that the reanalysis process simply reflects problems with *integration* rather than *prediction*, i.e. that it is simply more difficult to integrate a suffix which

has not been primed by its tone. Consequently, one could argue that no prediction is generated upon hearing the stem tone, but rather that the listener waits until all information is available before analysing the utterance as mismatching and that this is reflected in longer response times and late ERP effects.

Definitive evidence that prediction has taken place can only be found by isolating responses (behavioural or neurophysiological) that are made *before* the predicted constituent has been heard or seen. In the case of word accents, this means finding effects of prediction at the stem and signs that the suffix has been pre-activated by the tone.

Investigating neural correlates of predictive tonal cues

One of the first neurophysiological markers to suggest that word accents are used predictively was an ERP component found for Accent 2 (e.g. Roll et al., 2010, 2013) but not for Accent 1 stems (the positive-going 'P2'). The same component has been found for high left-edge boundary tones signalling main clause structures in Swedish (Roll, Horne & Lindgren, 2011).

Interestingly, the Roll et al. (2011) study on initial boundary tones also found that while the high tone functioned as a facilitating cue to upcoming main clause word order for the listener, it did not inhibit the processing of unexpected subordinate clause word order. Similarly, Accent 2 stems have been found to not inhibit the processing of mismatching Accent 1-associated suffixes as much as mismatching Accent 1 stems inhibit the processing of Accent 2-associated suffixes (Roll et al., 2010, 2013; Söderström et al., 2012). These findings have led to the suggestion that both Accent 1 and low left-edge boundary tones actually constitute more strongly constraining "micro-contexts" as compared to Accent 2 and high left-edge boundary tones. As has already been mentioned, Accent 2 stems can be argued to activate more word forms than Accent 1 stems – including both suffixed forms and compounds. From this, it follows that Accent 2 stems can be viewed as less constraining contexts. In a similar way, low left-edge boundary tones can be seen as more strongly constraining contexts since they signal only subordinate clause structure (Roll et al., 2011). This account is supported by findings that unexpected items encountered in strongly constraining contexts are associated with greater processing costs, possibly reflecting prediction revisions (Federmeier, Wlotko, De Ochoa-Dewald & Kutas, 2007). Furthermore, processing costs are also increased in contexts where fewer completions are available, as compared to those with more possibilities (Wlotko & Federmeier, 2012). In short, having excessively many options – as is the case for Accent 2 stems – makes prediction difficult.

Based on these assumptions, the ERP difference between Accent 1 and 2 stems has subsequently been reanalysed as a *negatively* charged effect for low left-edge boundary tones and Accent 1 stems rather than as positivity for high left-edge boundary tones and Accent 2. There are many well-known ERP negativities that have been associated with various types of anticipatory processing that occur before a reaction or the presentation of a feedback stimulus, such as the contingent negative variation (CNV, Walter, Cooper, Aldridge, McCallum & Winter, 1964), the readiness potential (RP, Kornhuber & Deecke, 1965) and the stimulus-preceding negativity (SPN, Damen & Brunia, 1987). However, examples in the literature of investigations into actual "pre-stimulus" pre-activation mechanisms in natural language processing are rare. One magnetoencephalography (MEG) study (Dikker & Pylkkänen, 2013) found evidence that picture primes can lead to the pre-activation of written nouns and another influential study (deLong, Urbach & Kutas, 2005) took advantage of the N400 component to show that listeners form graded predictions about upcoming items and specific predictions about specific phonological word forms (such as the English indefinite article *a*/*an*). In light of this, the Accent 1 negativity could be an important tool to further our understanding of the way in which linguistic material can be pre-activated in sufficiently constraining within-word micro-contexts.

Stem negativities as indices of suffix pre-activation?

Roll et al. (submitted) is the first study to specifically shed light on the Accent 1 stem ERP negativity. It was suggested that the strongly predictive status of Accent 1 is reflected in this stem negativity, which in turn is thought to index the pre-activation of suffix memory traces.

A recent ERP study (Söderström et al., submitted) set out to investigate the predictive

functions of Accent 1 and 2 stems more closely. The experiment in this study involved two methodological novelties. Firstly, it made use of pseudo-nouns – or rather pseudo-stems connected to either Accent 1- or 2-associated singular or plural suffixes – such as *tväk* (tfɛ:k) embedded in carrier sentences to test the hypothesis that the tone-suffix association still exists in the absence of lexical information on the stem. Secondly, it was the first to include a condition in which some suffixes were replaced with coughing sounds. This cough condition is important, as it makes it possible to directly test any effects of suffix pre-activation in the absence of the relevant prediction feedback stimulus (the suffix). As in previous studies, the participants' task was to determine as quickly as possible whether the word was in the singular or plural.

The study revealed a P600 effect for mismatching suffixes, suggesting that both mismatch combinations elicited reanalysis and reprocessing. It therefore seems likely that both word accents generate predictions which can be disconfirmed by mismatching suffixes. The P600 effect was preceded by a left-anterior negativity (LAN), which is thought to reflect morphological processing and the activation of memory traces of e.g. affixes which have not been properly primed (Pulvermüller & Shtyrov, 2003).

As regards the cough condition, it was noted that there was no difference in response times between Accent 1 and 2 (792 and 798 ms respectively, measured from cough onset), but response accuracy was significantly higher for Accent 1 words (87.8%) compared to Accent 2 words (72.0%). On its own, this suggests that Accent 1 stems cue their suffixes more strongly, but does not necessarily mean that the suffix was pre-activated. However, evidence of suffix pre-activation was found in a correlation between the amplitude of the Accent 1 stem ERP negativity and response accuracy: participants who displayed a larger negativity for Accent 1 stems also showed higher response accuracy in the cough condition. Furthermore, the scalp distribution of the Accent 1 stem negativity displayed similarities to a negativity found for suffixes compared to coughs (LAN), which suggests that suffix pre-activation and processing is indeed present before the suffix has even been perceived. One potential candidate for the brain area underpinning this suffix pre-activation mechanism is Brodmann area 47, which is thought to be involved in e.g. morphological processing (Roll et al., submitted). This is suggestive of an account according to which suffix processing is indeed initiated earlier for Accent 1 stems than for Accent 2 stems.

Conclusion

Results indicate that both word accents can be used predictively, but in different ways. Firstly, P600 effects have been found both for mismatching Accent 1 and Accent 2 words. Secondly, participants were relatively successful in restoring missing suffixes following both Accent 1 and Accent 2 stems (Söderström et al., submitted). Thirdly, the stem negativity points to a strong predictive role for Accent 1. The difference in the predictive status of the word accents thus seems to be based on Accent 1 generating stronger predictions for suffixes while Accent 2 stems generate weaker suffix predictions. The strong predictions allow listeners to commit to the ending of a word more strongly upon hearing an Accent 1 stem, as compared to Accent 2 stems.

Acknowledgements

This work was supported by the Swedish Research Council (grant number 2011-2284) and Knut and Alice Wallenberg Foundation (grant number 2014.0139).

References

Bar M (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7): 280-289.

Bruce G (1977). *Swedish word accents in sentence perspective*. Lund: Gleerups.

Damen EJP & Brunia CHM (1987). Changes in heart rate and slow brain potentials related to motor preparation and stimulus anticipation in a time estimation task. *Psychophysiology*, 24(6): 700-713.

deLong KA, Urbach TP & Kutas M (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8): 1117-1121.

Dikker S & Pylkkänen L (2013). Predicting language: MEG evidence for lexical preactivation. *Brain & Language,* 127: 55-64.

Federmeier KD, Wlotko EW, De Ochoa-Dewald E & Kutas M (2007). Multiple effects of sentential constraint on word processing. *Brain Research,* 1146: 75-84.

Friston K (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience,* 11: 127-138.

Kornhuber HH & Deecke L (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv*, 284: 1-17.

Pulvermüller F & Shtyrov Y (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *NeuroImage*, 20: 159-172.

Riad T (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, 122: 1352-1379.

Rischel J (1963). Morphemic tone and word tone in Eastern Norwegian. *Phonetica,* 10: 154-164.

Roll M, Horne M & Lindgren M (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain Research*, 1330: 114-123.

Roll M, Horne M & Lindgren M (2011). Activating without inhibiting: Left-edge boundary tones and syntactic processing. *Journal of Cognitive Neuroscience*, 23: 1170-1179.

Roll M, Söderström P & Horne M (2013). Word-stem tones cue suffixes in the brain. *Brain Research*, 1520: 116-120.

Roll M, Söderström P, Mannfolk P, Shtyrov Y, Johansson M, van Westen D & Horne M (submitted). Word tones cueing morphosyntactic structure: neuroanatomical substrates and activation time course assessed by EEG and fMRI.

Skipper J (2015). The NOLB model: a model of the natural organization of language in the brain. In: Willems, R, ed, *Cognitive Neuroscience of Natural Language Use*. United Kingdom: Cambridge University Press, 101-134.

Söderström P, Roll M & Horne M (2012). Processing morphologically conditioned word accents. *Mental Lexicon*, 7: 77-89.

Söderström P, Horne M & Roll M (submitted). Stem tones pre-activate suffixes in the brain.

van Boxtel GJM & Böcker KBE (2004). Cortical measures of anticipation. *Journal of Psychophysiology,* 18: 61-76.

van de Meerendonk N, Kolk HHJ, Vissers CTWM & Chwilla DJ (2008). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*, 22(1): 67–82.

Walter WG, Cooper R, Aldridge VJ, McCallum WC & Winter AL (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203: 380-384.

Wlotko EW & Federmeier KD (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, 62: 356-366.

Foreign accent: influences of the sound system of Serbian on the production of Swedish L2

Mechtild Tronnier[1] and Elisabeth Zetterholm[2]
[1] Centre for Languages and Literature, Lund University, Lund, Sweden
[2] Department of Language Education, Stockholm University, Stockholm, Sweden

Abstract

Among the second language learners of Swedish from the group of first language speakers (L1) of the one of the standard languages that formed the pluricentric language Serbo-Croatian, Serbian currently seems to be most frequent L1. Based on recordings of L2-speech produced by two Serbian L1-speakers, living in Sweden, this contribution presents typical pronunciation features in L2-Swedish, in order to provide teachers with a better understanding of foreign accent characteristics when teaching Swedish as a second language.

Introduction

Due to the Balkan conflict in the mid-90s, there has been a flow of L1-speakers of any of the languages from former Yugoslavia migrating to other European countries. This flow of migration culminated in Sweden in 1994 (www.migrationsverket.se). According to the same source, however, some migration is still occurring, as their statistics show that ca. 500 applications for residence permits were handed in during the first four months of the year 2015. The applicants came exclusively from the Republic of Serbia.

In a survey obtained in 2011 (Tronnier & Zetterholm 2011) it was shown that L1-speakers of Serbian/Bosnian/Croatian were regularly found in the classrooms of Swedish as a second language, although they were greatly outnumbered by L1-speakers of e.g. Arabic and Somali.

Serbian was one of the standard languages of the pluricentric language Serbo-Croatian and is nowadays mainly spoken by the Serbs in what is now Serbia, Bosnia and Herzegovina and Montenegro. Croatian and Bosnian are the two other standard languages which formed Serbo-Croatian. The languages Serbian, Bosnian, and Croatian are all three of the Western South Slavic type. They differ slightly, but are mutually intelligible languages.

Among the students of L2-Swedish with any of the standard languages of the pluricentric Serbo-Croatian language as their L1, Serbian speakers are currently the most common. Therefore an analysis of foreign accent features observed for speakers with Swedish as their L2 and Serbian as their L1 will be presented in this contribution.

In the following section, the sound inventory of the Serbian language will be presented in a contrastive way, i.e. in comparison to the Swedish sound inventory. In addition, common pronunciation features in Swedish-L2 produced by speakers with Serbian as their L1 will be presented.

The sound system of Serbian in comparison with Swedish

Vowels

The Serbian vowel system comprises five phonemes /i e a o u/, whereas nine vowel phonemes can be found in Swedish /i y e ɛ ø a o u ʉ/, each of those found in long and short forms (Bruce, 2014). Basically, all vowels in Serbian are also present in the Swedish vowel system, but not the reverse. Vowels present in Swedish but not in Serbian include the front rounded long and short vowels /y ø/ and the central vowel /ʉ/. Furthermore, the distinction between the unrounded front close-mid and open-mid vowels /e ɛ / is not made in Serbian. Serbian presents us with only one vowel phoneme in that region: /e/, which – with some exceptions – has the quality of the open-mid vowel [ɛ] (Morén, 2006; Petrović, 2001). In addition, the back mid vowel /o/ is realized as an open-mid vowel, resulting in [ɔ].

In most cases, the short vowels differ in quality from the corresponding long vowels in Swedish in that the short vowels tend to be more

close. Variation in vowel quantity in Serbian, however, is related to the tonal accent and does not affect vowel quality.

Consonants

There is a large overlap in the consonantal inventory between Serbian and Swedish. In the case of voiceless fricatives, /ɕ/ and /ɧ/ occur in Swedish, and are not part of the Serbian consonantal inventory. Serbian, however, includes the fricatives /ʃ/ and /x/, which are very close to those two in Swedish mentioned above.

Different labels are given for the back fricative in Serbian: in some sources it is named /h/ (Morén 2006; Petrović, 2001) and in some sources it is named /x/ (Speech accent archive; Garlén, 1988). No matter the label, there seems to be an agreement that the realisation is close to a velar fricative, namely [x]. Where there are two voiceless fricative phonemes in Swedish – namely /h/ and /ɧ/ – only one can be found in Serbian.

Furthermore, there is a discrepancy between nasals: Swedish presents a velar nasal /ŋ/, which does not occur in Serbian, which on the other hand presents a palatal nasal /ɲ/. The nasal /n/, however assimilates to [ŋ] in Serbian when placed prior to velar consonants (Garlén, 1988).

Two lateral consonants can be found in Serbian /l/ and /ʎ/, whereas only one – /l/ – is considered a phoneme in Swedish. According to Gick et al. (2006), /l/ in Serbian is velarized: [ɬ].

The consonants /r/ and /l/ take the role of the nucleus of the syllable, when they occur between consonants (Petrović, 2001).

Prosody and syllable structure

Serbian – similarly to Swedish – has flexible stress placement. In both languages, placement is to a large extent based on the morphology of the word. This implies that the speaker would need to know about the meaning and grammatical role – and probably history – of the morphemes assembled in a word to place the stress on the correct syllable.

In Swedish, quantity contrast appears on stressed syllables only (Bruce, 2012), while in Serbian it also occurs on unstressed syllables. Quantity contrast in Serbian is based on variation in vowel length, whereas in Swedish a complementary length variation materializes in most dialects. In the case of complementary length in Swedish, long vowels are coupled with a short or no consonant in the rhyme of the stressed syllable and short vowels are coupled with long consonants (geminates) or consonant clusters in the rhyme of the stressed syllable. (Bruce, ibid.)

Tonal word accents occur in both languages, Swedish (Bruce, ibid.) and Serbian (Petrović, 2001). Despite the fact that the question of word accent use across tone accent languages is very interesting, no detailed analysis of this matter is presented in this contribution. The tonal shape of word accents varies a lot across the Swedish dialects – with some dialects not making any contrast. In addition, experience shows that incorrect application of the word accents does not lead to miscommunication in the first place. A question addressing word accent use across tone accent languages deserves a deeper analysis.

Consonant clusters occur in both languages and the number of consonants allowed in syllable initial and final position in both languages is two-three consonants. Nonetheless, a larger range of consonant clusters types are allowed in syllable initial position in Serbian than in Swedish.

Analysis of foreign accent features

Material and subjects

For the present study recordings were made of two female speakers with Serbian as their first language speaking L2-Swedish. The two speakers both lived in the south of Sweden and were both enrolled in classes for Swedish as a second language when the recordings were made. They had a good command of conversational Swedish and also reported on some proficiency in English. The recordings consist of read speech of Swedish sentences and a short text, and elicited spontaneous speech, namely the description of a picture story. The reading material was constructed so that all the Swedish segmental and prosodic phonemic inventory was present. This also includes some minimal word pairs, compound words and words with typical Swedish consonant clusters.

The foreign accent analysis was then carried out auditorily: instances of pronunciations which

Table 1. Overview over frequently diverging vowel pronunciation in L2-Swedish produced by L1-speakers of Serbian.

Phoneme	pronunciation L1-Swedish	example	pronunciation L2-Swedish	summary
1. /iː/	[iː]	*vit* [viːt] "white"	[ɪ]: *[vɪtʰ]	/iː/ ([iː]) → [ɪ]
2. /y/	[ʏ]	*mycket* [mʏkːɛ(t)] " a lot"	[ɪ]: *[mɪkːɛtʰ]	/y/ ([ʏ]) → [ɪ]
3. /eː/	[eː]	*spela* [speːla] "to play"	[ɛː]: *[spɛːla]	/eː/ ([eː]) → [ɛː]
4. /øː/	[øː]/[œː]	*röda* [røːda] "red"	[ɔː]: *[rɔːda]	/øː/ ([øː]) → [ɔː]
5. /ø/	[œ]	*höst* [hœst] "fall"	[ɔː]: *[hɔːst]	/øː/ ([œ]) → [ɔ]
6. /oː/	[oː]	*håret* [hoːrɛt] "hair"	[ɔː]: *[hɔːrɛtʰ]	/oː/ ([oː]) → [ɔː]
7. /ʉː/	[ʉ̟ː]	*hus* [hʉ̟ːs] "the house"	[uː]/[ʊː]: *[hʊːs]	/ʉː/ ([ʉ̟ː]) → [uː]/[ʊː]
8. /aː/	[ɑː]	*svag* [svɑːg] "weak"	[aː]: *[svaːg̊]	/aː/ ([ɑː]) → [aː]

did not match with what was expected for Swedish were marked and transcribed. Major and clearly noticeable differences and recurring differences in pronunciation observed in the recorded material are presented in the following section of this contribution.

Vowels

The overview in Table 1 shows the most frequent divergence of vowel production in L2-Swedish. It seems that the diversity of the vocalic inventory in Swedish was accommodated within the five-vowel system of Serbian. In that respect the different levels of mid vowels in Swedish, which consists of vowels of mid close quality ([e o]) and mid open quality ([ɛ ɔ]) were not distinguished by the L2-speakers. In most of the cases, the mid vowels were realised as open mid vowels, (cf. example 3. and 6. in Table 1).

Moreover, the difference in quality between the short and long variant of /a/ in Swedish was not produced. Instead, the long open vowel /aː/ – which is realised as a more back vowel [ɑː] in Swedish – was produced by the L2-speakers with the same quality as the short open vowel /a/ – which is [a] – but with a longer duration, resulting in [aː] (example 8. in Table 1).

The rounded front vowels are usually one of the obstacles for L2-learners of Swedish. They were also affected by the L2-speech of the speakers we investigated, with Serbian as their L1. Thus, the rounded close front vowel /y/ was produced unrounded, resulting in [i] (example 2. in Table 1). The rounded midvowel /ø/ – both long and short, however, was not similarly altered by making use of the opposite lip articulation, but the tongue position was moved to a back articulation instead, resulting in [ɔ] – in the long and short members of the category (examples 4. and 5. in Table 1). In this case spelling conventions might have played a role, as the letter <ö> – representing the phoneme /ø/ – is rather unusual in many languages. Serbian, on the other hand, is written with the Cyrillic alphabet so that the Roman alphabet should not have interfered necessarily. Nonetheless, the subjects reported on some knowledge of English, which is written in Roman letters.

The pronunciation of /ʉ/ might also be influenced by the spelling, as its corresponding letter is <u>. Both, the long and the short members of /ʉ/ were realised with a quality of [u] or [ʊ] by the L2-speakers (example 7. in Table 1). The letter <o> also contributes to pronunciation errors. In some cases the corresponding phoneme is /o/ and in other cases it is /u/, which leads the L2-speakers of Swedish to a preferred pronunciation of [ɔ], as in the example *stol* "chair" [stuːl] which was realised as *[stɔːl].

There was a general tendency for the Swedish long vowels to be pronounced slightly more open by the L2-speakers than what would be expected in Swedish (cf. examples 1., 3., 6. and 7. in Table 1).

Consonants

In all cases, diverging pronunciation of consonants was based on a micro-level, i.e. only single features deviated from native pronunciation. There was a general tendency of word final devoicing for voiced stops: e.g. *svag "weak"* [svɑːg] is pronounced as *[svaːg̊] in the L2-Swedish studied. Furthermore, voiceless final stops appeared with a clear aspiration, even when the final syllable should be unstressed, which is unusual for L1-Swedish. In that way,

rummet ['rʊmːɛt] "the room" was pronounced as *['rʊmːɛtʰ]. On the other hand, aspiration was lacking where expected, namely when a voiceless stop introduced a stressed syllable, as in *pappa* "daddy" ['pʰapːa] which was pronounced *['papa] (Fig. 1).

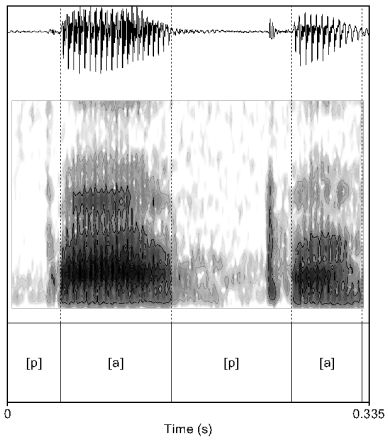

*Figure 1. One of the speakers producing the word **pappa** without the aspirated stop, which usually introduces the initial – stressed – syllable.*

The velar nasal [ŋ] occurs in Serbian as a contextual allophone of /n/, when another velar consonant follows. The Serbian L1-speakers addressed the Swedish phoneme /ŋ/ in a similar way: the Swedish spelling is <ng> and that might delude the reader to think that it should be pronounced as the sequence [ŋg], a voiced velar stop, which is preceded by a nasal, assimilated to the same place of articulation. The word which exemplifies such transfer is *många* "many" [mɔŋa], which was realized as *[mɔŋga]. Such a realization also occurred in the section of spontaneous speech and was therefore not only a result of reading. However, as adult learners tend to learn pronunciation factors in relationship to the written language than by listening to spoken language only, an assumption about pronunciation might have been made from spelling that also affects spontaneous speech.

It has been mentioned above that devoicing and aspiration of final stops was observed for the L2-speakers of Swedish. These components of foreign accent were also added to a word final velar nasal /ŋ/. In that way, a voiceless and aspirated homorganic stop was added to /ŋ/. For example the word *lång* "long" [lɔŋ] (/loŋ/) was pronounced as *[lɔŋkʰ] (Fig. 2).

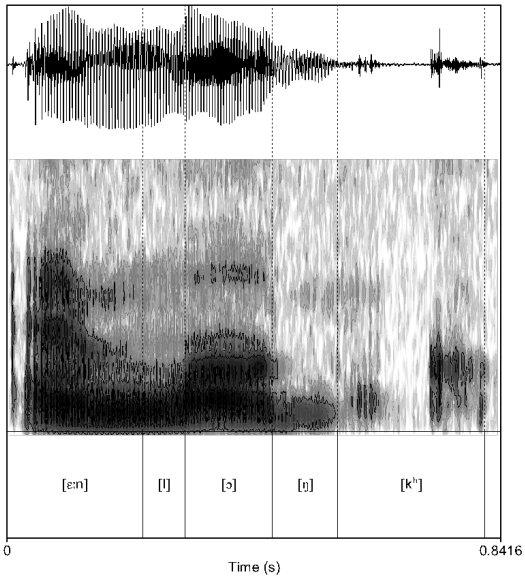

*Figure 2. One of the speakers producing the word **lång**, adding a voiceless aspirated stop after the velar nasal.*

In the southern variety of Swedish the lateral consonant /l/ is realized as [l]. Serbian has two laterals, one of which is palatal, /ʎ/ ([ʎ]) and the other one, commonly labeled as /l/, which usually is velarized: [ɬ]. This feature occurred frequently in L2-Swedish, e.g. the word *flygplatsen* "airport" [flyːgplatsən] was pronounced as *[fɬɪːgpɬatsən].

The realization of /r/ as a trill [r] is possible in both Swedish and Serbian. In Swedish, however, it is realized as such when introducing a stressed syllable rather than in other cases, where variants of /r/ with a weaker articulation is preferred. The L2-speakers also applied the trill in syllable final position, and moreover in the final positions of unstressed syllables. This results in an appearance of /r/ which is unusually salient for Swedish. The sequence of words *tycker många* "many (people) find/think (that)" [tʰʏkːɛ(ɹ) mɔŋa] was thus realised as *[tɪkɛr mɔŋga].

Consonant clusters and the voiceless Swedish fricatives [ɧ] and [ɕ], which are usually difficult for L2-speakers, were not compromised.

Prosody and syllable structure

When it comes to placement of stress and vowel quantity, errors were made only occasionally. In addition, consonant clusters did not seem to cause any difficulties either.

What was rather striking was that when a short vowel was produced correctly, the following complementary geminates appear to be too short, as in *pappa* "daddy" ['pʰapːa], which was pronounced as *['papa] (see Fig 1). Such pronunciation compromises the linking of the syllables in that a juncture is perceived in L2-speech between the vowel in the stressed syllable and the following consonant, the latter seemingly to introduce the post stressed syllable as a whole. Such a clear break between the syllables does not arise in L1-speech, where a part of the geminated consonant concludes the stressed syllable, and some other part of the geminate introduces the following syllable. With the juncture placed in a variant position, the flow of speech receives a rhythmic structure, which is unexpected.

A similar observation can be made about the use of /r/ in intervocalic position. As has been mentioned above, the pronunciation of /r/ as a trill in syllable final position makes it unusually salient. One further observation was that such pronunciation was realised in that way, that the syllable final – and morpheme or word final – /r/ seemed to be shifted into the syllable initial position of the following syllable, so that its association was moved from coda position to onset position. The sequence of *för att* "so that" [fœɹ.at], which resulted in *[fɔ.rat] serves as an example, where the juncture between the syllables shifted to the place before /r/.

As pointed out above, the velar nasal /ŋ/ was almost always produced as [ŋg]. In intervocalic position the homorganic stop was then associated with the following syllable – within a word and also across word boundaries. When found at a word boundary, such a sequence obscures the identification of running speech as the appropriate word after /ŋ/ does not start with [g]. One example is: *sprang över* "ran across" [spraŋ.œː.vəɹ], which was pronounced as *[spraŋ.g̊œː.vər] in L2-Swedish.

Other observations

Rules of assimilation – which feature does have an influence and whether it applies either backwards or forward – are a language specific characters. In Swedish, voiceless consonants usually have a stronger influence on voiced ones with which they are in immediate contact in either the same consonant clusters or across syllable and word boundaries. This often leads to the devoicing of the voiced consonants – although sometimes only partially. It was observed, that L2-speakers produced the reverse: i.e. voiced consonants affected adjacent voiceless ones, which in turn became voiced. Thus, the word *glänste* "shone" [glɛnstə] – where the /n/ may be partially devoiced– was produced as [glɛnzdə̬], where the fricative became voiced, [z] – a sound that does not occur in Swedish at all. Such realisation of assimilation, however, agrees with the rules in Serbian – and for that matter also in other Slavic languages. In addition, the stop that followed was produced – although not voiced – with a weak burst.

Conclusions

Some of the observations from L2-speech described above have a stronger influence on intelligibility of communication than others. The factors which might make communication more difficult are presumably those which interfere with the rhythmic structure. In that way, the absence of germination, which leads to a coda consonant been shifted into onset position, the introduction of an extra consonant – including a consonant taking an onset position where no consonant is expected – and the strong articulation of /r/ – which is rare in final position and therefore likely to be apprehended as syllable onset – are seemingly effective modifications. More research based on an experimental procedure is needed to verify such assumptions.

Final devoicing of consonants, including aspiration, lack of syllable initial aspiration, introduction of a velar consonant after a velar nasal – unless preceding a vowel –, assimilatory voicing of voiceless consonants and most of the modifications of vowel quality made by the L2-speakers probably sound somewhat atypical rather then lead to problems in communication. A systematic analysis based on recognition tests would be beneficial to shed a clearer light on this matter.

References

Bruce, G. (2010). *Vår fonetiska geografi.* Lund: Studentlitteratur.

Bruce, G. (2012). *Allmän och svensk prosodi.* Lund: Studentlitteratur.

Garlén, C. (1988). *Svenskans fonologi*. Lund: Studentlitteratur.

Gick, B., F. Campbell, S. Oh and L. Tamburri-Watt. (2006). Toward universals in the gestural

organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34:1, 49-72.

Morén, B. (2006) Consonant-Vowel Interactions in Serbian: Features, Representations and Constraint Interactions. *Lingua* 116, 1198-1244.

Petrović, D (2001). Languages in Contact: Standard Serbian phonology in an urban setting. *Int'l. J. Soc. Lang.* 151: 19-40. Walter de Gruyter.

Speech accent archive, http://accent.gmu.edu, retrieved 20150428.

Swedish Migration Board, http://www.migrationsverket.se, retrieved 20150505.

Tronnier, M & Zetterholm, E (2011). New Foreign Accents in Swedish. *Proc. 17th Int. Congress of Phonetic Sciences (ICPhS2011)*, Hong Kong, 2018–2021.

Halfway to Estuary English with H. G. Wells (1866-1946)

Sidney A. J. Wood

Abstract

The speech of seven Kentish informants born in the closing decades of the 19th century (recorded by the Survey of English Dialects*), and of their contemporary, the author H. G. Wells (also Kentish), is reviewed with respect to six sound changes already present in varying combinations in their Kentish speech, in order to elucidate the routes and timing for how they were being spread. The recordings have caught moments in the period 1860-95 where a new accent (Estuary English) was spreading across rural Kent from the towns along the shore of the Thames estuary, each informant exhibiting an individual mixture of new pronunciations and earlier Kentish pronunciations. Similar sound changes were occurring throughout the home counties, but their precise timing, progress and routes were not studied for this article. The changes examined here are loss of rhoticity,* TRAP *shifted away from* DRESS, THOUGHT *modified from* [ɔː] *to* [oː], LOT *from* [a] *to* [ɔ], PRICE *from* [ʌi] *to* [ai], *and* MOUTH *from* [ɛʉ] *to* [æɒ]. *The results indicate that the sound changes started at different times, and spread through Kent over several generations, each starting from the north (London and the estuary coast), and ending in the east and south. Each change appears to have started spreading to neighbouring rural areas within a generation of appearing in estuary towns. The earliest of the six changes was* PRICE*, estimated appearing around 1800 in estuary towns (acquired by seven informants); then* THOUGHT *(acquired by five);* MOUTH *and rhoticity loss (four each);* TRAP *(three); and the most recent was* LOT *(two).*

Introduction

This talk has two topics: (i) Estuary English (EE), the new accent that spread through south east England in the 19th century, and (ii) the speech of biologist and author H. G. Wells, an example of someone who had acquired part of this new accent.

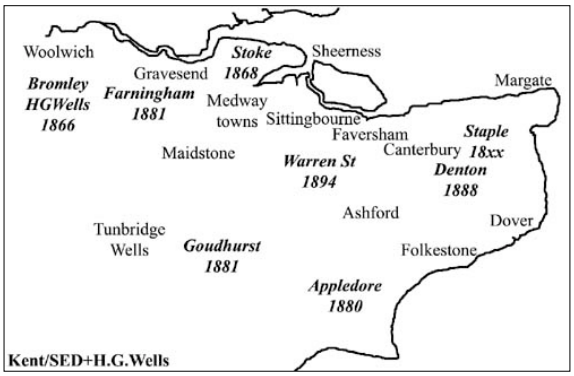

Figure 1. The locations and birth years of the seven SED informants and H G Wells (in italics), and various towns.

The recordings were made by the Survey of English Dialects (SED) (Robinson) in the 1950s, explicitly targeting rural areas in order to find more surviving dialect features. They are available online in MP3 format from the British Library. Figure 1 shows how the SED informants were distributed across rural Kent, together with neighbouring urban areas. They were born in the interval 1860-1895, all were male, and had such varied occupations as blacksmith, coal miner, farmer, groom and traction engine driver. The recording of H. G. Wells is from a BBC broadcast (BBC Archive 1931). Wells was born in 1866 at Bromley, at that time a large village growing into a substantial urban area, later into a suburb, and now the centre of a London Borough. His speech is comparable to that of the contemporary SED informants, except it is on the highest end of the regional sociolect scale (the SED informants all have popular accents).

The following phonetic character substitutions are made for this article: (i) the character **a** (offici-ally *open front* cardinal 4) is used for vacant *open central*, and (ii) the character **æ** for unshifted TRAP (officially *lowered half-open front*) is also used for shifted TRAP (officially *open front* **a** adjacent to cardinal 4).

Expressions like FOOT, STRUT, TRAP, BATH are keywords used by Wells (1982) for lexical sets that took part in various sound changes. They are more convenient than phonemes for studies of dialect or sound change, where actual pronunciations are being modified and phoneme systems revised.

Lossy MP3 compression had degraded the online sound files so that formant analysis by LPC formant tracking in Praat (Boersma & Weenink 2014) yielded inconsistent results. Formants were consequently identified and measured on FFT slices from narrow band spectrograms. Only fully stressed exemplars of vowels taken from focally accented syllables were analysed in order to exclude variation due to vowel reduction. Formants were measured at the moment where the vowel was least affected by adjacent consonants (to minimize measurement scatter caused by coarticulation effects), determined by observing CV and VC formant transitions on spectrograms. A typical five minute recording of spontaneous speech might yield 5 or more focally accented instances of the most frequent vowel phonemes, but just a few or even none of the least frequent.

Background

Southern British English

Southern British English (SBE) is the dialect of English spoken in England south of a line roughly between the Wash in the east and the Severn Estuary in the west (Wales 2006, Map 1.6). It is distinguished from Northern English by two sound changes (that occurred in the 15th-17th centuries): the FOOT-STRUT split and the TRAP-BATH split. In SBE, words in the FOOT and STRUT classes and the TRAP and BATH classes are pronounced with four different vowel phonemes (today roughly [ʊ] and [a], and [æ] and [ɑ:] in the home counties).

Four regiolects are recognised within SBE: East Anglian, London, Home Counties and Western. Within each regiolect, there is social variation along a scale from popular to standard sociolects (Gimson 1962 §6.92, Wells 1982 §1.1.5).

Estuary English

The expression *Estuary English* became well known following Rosewarne (1984), unfortunately often misunderstood, or worse still vulgarized as Received Pronunciation (RP) (Jones 1918, Gimson 1962, Wells 1982) modified towards East End Cockney.

During the 18th and 19th centuries, the popular London accent had been brought by Londoners migrating along both shores of the Thames Estuary, occasioning the sound changes of EE in the local population. That is the story we were told at school by our English master in 1950. We spoke Estuary English, our teacher said, we were bilingual, switching as we came in through the school doorway. Translated into linguistic terms, we were diglossic with two sociolects: popular EE on the playing field and standard EE inside school. The staff had graduated from a fair spectrum of British universities of the 1920s-1940s, none spoke RP, a few were from the north and spoke Northern English, but most were from the south east and became our model for standard EE.

However, similar changes were occurring throughout the Home Counties and today the accent is fairly uniform right across this area, and not just around the Thames estuary.

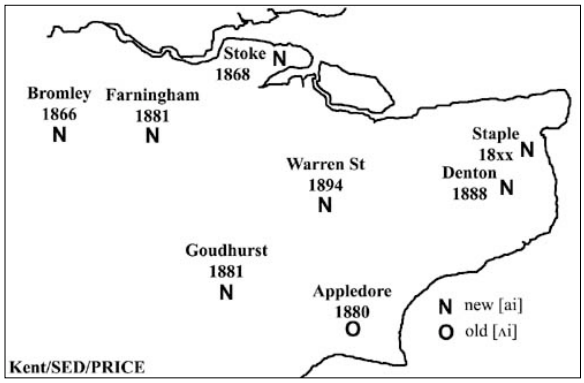

Figure 2. The distributions of PRICE pronunciations in rural Kent. This sound change was almost complete in Kent by 1860-95, the earlier pronunciation still occurring in the SE.

Parish & Shaw (1888 p. vii) gave an introductory account of how the London accent was being spread via the estuary towns, although, sadly, it reveals more about these authors' attitude towards popular speech than about their linguistic skill:

> "The purity of the dialect diminishes in proportion to the proximity to London of the district in which it is spoken. It may be said that the dialectal sewage of the Metropolis finds its way down the river and is deposited on the southern bank of the Thames, as far as the limits of Gravesend-Reach, whence it seems to over-flow and saturate the neighbouring district."

Calling the new accent *metropolitan sewage* is a measure of how it was regarded. Similarly, H G Wells, who never modified his EE accent to RP, was called a *Cockney upstart*. One discriminating feature between EE and RP is MOUTH. EE, like popular London, has [æɒ] while RP has [au]. The late Edward Heath (1916-2005, UK prime minister 1970-74, born near Margate at Broadstairs) partially modified his native EE towards RP, but retained the [æɒ]-like MOUTH unchanged. He was satirized for that by *Monty Python's Flying Circus* (1971) in a sketch *Teach Yourself Heath*, a merci-less example of how EE was still regarded fifty years ago. The [æɒ]-like MOUTH was one of the shibboleths that immediately gave away anyone who was not speaking RP.

The SED recordings have caught new pronun-ciations for EE spreading from the estuary towns to rural Kent in the period 1860-95, the informants exhibiting varying mixtures of new features and earlier Kentish features, which permits the progress of this new accent to be evaluated (Wood, forthcoming).

These new pronunciations represent six sound changes that were present in varying combinations in the speech of these informants: PRICE modified from [ʌi] to [ai], THOUGHT modified from [ɔː] to [oː], MOUTH modified from [ɛʉ] to [æɒ], rhoticity lost (/r/ no longer pronounced in syllable codas), [æ] for TRAP shifted away from DRESS, and LOT modified from [a] to [ɔ].

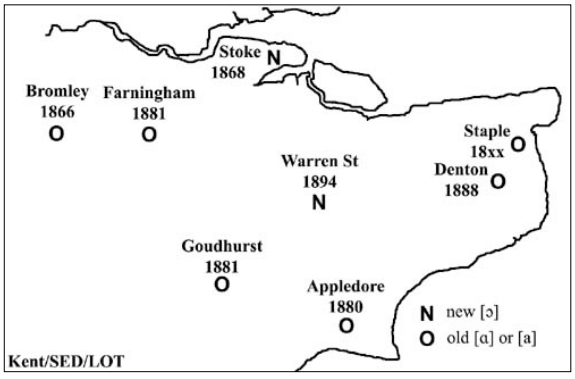

Figure 3. The distributions of LOT pronunciations. This sound change had barely started and was the most recent of the six.

Figure 2 records the distributions of PRICE pronunciations, showing that six SED informants scattered almost over all the county, and H G Wells, had already acquired the new [ai]-like timbre for PRICE, suggesting this sound change had commenced much earlier in the estuary towns and was almost complete in the county by 1860-95.

For comparison, Figure 3 records the LOT pro-nunciations, showing that only two informants had acquired the new [ɔ]-like timbre, which was conse-quently one of the more recent changes, starting to spread in the 1860s and established in the Medway towns since the 1850s.

There were three different combinations of THOUGHT and LOT forms among the informants, summarized in Table 1: new THOUGHT and new LOT, new THOUGHT and earlier LOT, and earlier THOUGHT and earlier LOT. Oddly, the precise timbre of the earlier [a]-like LOT varied between the informants, from an extremely bright [a] to a darker [ɑ]. This is possibly not so random as it might seem (although there are only eight informants). The darker [ɑ] for LOT is by the three informants who still have the earlier THOUGHT and earlier LOT, while the brighter [a] for LOT is by the three informants who have the new THOUGHT together with the earlier LOT.

Combination	Informants
New THOUGHT [oː] New LOT [ɔ]	*Stoke 1868* *Warren St 1894*
New THOUGHT [oː] Old LOT [a]	*Appledore 1880* *Farningham 1881* *Staple 18xx*
Old THOUGHT [ɔː] Old LOT [ɑ]	*Goudhurst 1881* *Denton 1888* H G Wells *Bromley 1866*

Table 1. Combinations of THOUGHT and LOT outcomes.

If PRICE was the earliest of these six sound changes, and LOT the most recent, a rough timetable can be established for each of the six sound changes by timing progress across the county, assuming roughly a generation to spread between each informant's location when there is no better information. In addition, other sound changes were complete, and yet others had not commenced. There was clearly a long succession of individual sound changes at regular intervals starting back in the 18th century and continuing into the 20th century:

- PRICE (Figure 2) had spread to most of the county by the 1880s and was acquired by 7 out of the 8 informants. Assuming three or four generations to spread, this points to this change appearing by 1800 on the estuary coast.
- THOUGHT had reached parts of East Kent by the 1880s, and had been acquired by 5 out of

the 8 informants. It had spread more than loss of rhoticity but less than PRICE, and had perhaps started in the 1820s or 1830s in the Medway Towns, later further west (H G Wells, *Bromley 1866*, had not acquired the new THOUGHT).

- MOUTH had reached the northern half of the county, from Bromley in the west (H G Wells) to *Staple 18xx* in the east and had been acquired by 2 informants and 1 partially and 1 transitionally; perhaps present in any of the estuary towns including Margate by the 1830s or 1840s
- Loss of rhoticity had reached the north-western half of Kent by 1894, 3 informants and 1 partially; similar to MOUTH, perhaps by the 1840s in estuary towns, while Ellis (1889) had evidence from Margate around the 1850s.
- TRAP had reached the central half of Kent, three informants, but excluding *Stoke 1868* and H G Wells (1866) at Bromley suggesting it started spreading to rural areas around the 1870s; perhaps in the Medway Towns by the 1850s.
- LOT (Figure 3) had reached only *Stoke 1868* and *Warren St 1894*, but not *Farningham 1881* or H G Wells at Bromley; perhaps in the Medway Towns by the 1850s.

Other, even earlier, sound changes were complete by 1860 and were shared by all the informants. For example, they all had the new [ao] for GOAT (the earlier pronunciation had been something like [ou]). A later sound change (after 1900) darkened BATH to [ɑː], but all these eight informants still had the earlier [aː].

Table 2. The acquisition of the six sound changes (ordered from the left by number acquired from most to fewest) by the seven SED informants and H G Wells (ordered from the top from west to east), showing completely acquired new changes (N), partially acquired (P), transitional (T), and older form still acquired (O).

Informant	PRICE	THOUGHT	TRAP	MOUTH	Nonrhotic	LOT
H G Wells	N	O	O	N	N	O
Farningham	N	N	N	N	N	O
Stoke	N	N	O	O	P	N
Warren St	N	N	N	P	N	N
Goudhurst	N	O	N	O	O	O
Staple	N	N	O	T	O	O
Denton	N	O	O	O	O	O
Appledore	O	N	O	O	O	O

Table 2 indicates two factors that determined how these sound changes spread through Kent: the younger informants tended to have acquired more sound changes, and the informants located nearer the estuary tended to have acquired more sound changes. Each sound change started from estuary towns and spread gradually towards East Kent and the SE.

The earlier Kentish accent

The main sources for the earlier accent in Kent are Ellis (1889) and Parish & Shaw (1888). Ellis rejected several districts because the earlier dialect was no longer spoken there, particularly the Hoo peninsula (opposite the Medway Towns, *Stoke* in Figure 1), the Isle of Sheppey (including Sheer-ness), and Thanet (including Margate, just north of *Staple* in Figure 1), all located along the estuary coast and close to estuary towns. Presumably that also meant EE was already becoming established along the whole length of the estuary coast by the 1870s, or a generation earlier, perhaps already by 1850.

Ellis reported the Kentish /r/ as a "burr", which in the 19th century usually meant uvular (Sweet 1892: 31). This is how I also perceive the SED Kent informants' /r/, and my own, a uvular continuant (not a trill or fricative). This was presumably not a spontaneous innovation of the 1880s but continued from the earlier pronunciation.

Ellis had heard fully rhotic speech in the towns of Tunbridge Wells and Maidstone, but gave no ages of informants or dates. Otherwise, /r/ had "*a tendency to degenerate into the ordinary English vocal r, a mere vowel (ə ɐ) … the form it retains in London*", i.e. a general tendency towards non-rhotic EE. In the coastal town of Margate, a student teacher's /r/ "*followed London use*" including "*euphonic insertion*", which I understand as meaning it was non-rhotic with linking and intrusive **r**. Of a student teacher at Charing (between Ashford and *Warren Street 1894*, Figure 1) Ellis noted "*with slight exceptions all recollection of the dialect seemed to have left her, the (r) was quite Cockney*". These student teachers were training at the esteemed Whitelands College in Chelsea (now Roehampton University) but there was no suggestion they were modifying their obvious EE tendencies towards RP.

Other pronunciations Ellis noted, transliterated into the IPA alphabet from his own palæotype transcriptions, are [ʌi] for PRICE and [ɛʊ] for MOUTH. He had heard that /ð/ was

pronounced [d], but failed to confirm it himself on Thanet or at Folkestone, accepting the word of his corres-pondents elsewhere. None of the SED Kent informants did it. It is preserved in the hundred verses of *Dick and Sal at Canterbury Fair*, written by John White Masters (1791-1873) of Sheldwich (a village near Faversham), published in Canterbury around 1820, and reproduced by Parish & Shaw as hard evidence of early 19th century rural Kentish speech. Masters was a horticulturist and pioneer of the Assam tea trade rather than a linguist. Here is one of the verses as published, in Masters' own modified spelling in order to simulate the actual pronunciation (the suggested narrow phonetic transcription is mine, writing ʀ for a uvular continuant and **a** for *open central*):

He sed dare was a teejus fair,
Dat lasted for a wick
And all de ploughmen dat went dare,
Must car dair shining-stick.

[i sed dɛʀ wə z ə ˈtiːʤə s fɛʀ
də ʔ ˈlaːstɪd fə ʀ ə wɪk
ə n ɔːw də ˈplɛɯmə n də ʔ wenʔ
dɛʀ mə s kɛʀ dɛʀ ˈʃʌinɪn stɪk]

All the Kent SED informants have glottal stops, vocalized /l/, and /h/-dropping, in common with people all over the country, so these are not obviously uniquely London features. I assume they are older. There is also a distributional difference. None of the SED Kent informants had intervocalic glottal stops, perhaps because instead they had /t/-voicing (laxing, lenition, flapping) between syllabics, especially when stress is weakened or absent. Some examples: *get home* [ged ˈəom] (*Appledore 1881*), *knock it on* [nɑk ɪd ˈɑn] (*Staple 18xx*), *at half past* [əd afˈpaːst] (*Goudhurst 1881*). With full focal stress, there is more likely to be an aspirated [tʰ].

H. G. Wells

Wells was born in 1866 at Bromley, just outside London. His parents kept a sports and chinaware shop in the village. His father was also a profes-sional county cricketer and his mother a servant.

Mugglestone (2003, p. 263) states that he "*shed the Cockney markings of his youth*", which probably meant he moved his accent up the sociolect scale to Standard EE (Mugglestone, p. 56, points out that C*ockney* was widely used in prescriptive rhetoric as a stereotype for any undesirable language). There are no signs of modification towards RP in his speech. In the 1931 broadcast, his voice is high pitched, but he speaks very clearly with no syllable contractions, there are no dropped /h/, no vocalized /l/, and very few glottal stops (none intervocalic). On the other hand there are unmodified bright [ʉ]-like GOOSE vowels and yod coalescence (both shunned in 19th and early 20th century RP), together with [æɒ] for MOUTH and [ao] for GOAT (that have never yet been recognized as RP, which has [aʊ] and [əʊ] respectively). These characteristics of his accent can be seen in Figures 4 and 5.

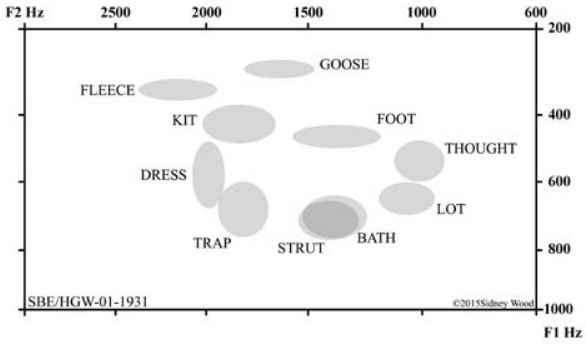

Figure 4. F1/F2 diagram for H. G. Wells' monophthongs, showing TRAP near DRESS, the same [a]-like timbre for STRUT and BATH (F2 around 1400Hz), the [a]-like LOT (F1>600Hz),the [ɔ:]-like thought F1 500-600Hz, and the bright [ʉ:]-like goose (F2>1500Hz).

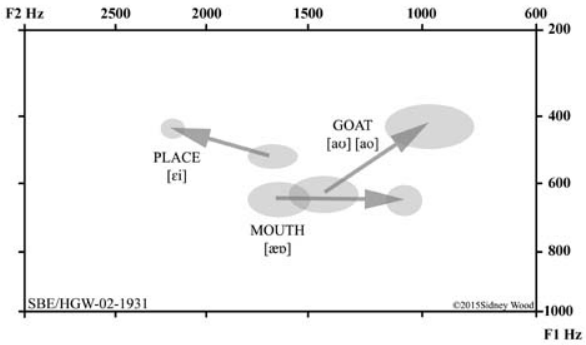

Figure 5. F1/F2 diagram for three diphthongs by H. G. Wells, all regional and none RP: [ɛi]-like PLACE (F1 starting from around 550Hz), [ao]-like GOAT (starting from F1>600Hz and F2 around 1400Hz), and [æɒ]-like MOUTH (F1>600Hz from start to end).

Table 2 records that he had acquired 3 of the 6 sound changes (he was non-rhotic, and he had [ai] for PRICE and [æɒ] for MOUTH). Otherwise he still had the earlier TRAP near DRESS, [ɔː]-like THOUGHT and [ɑ]-like LOT. For comparison, his nearest SED neighbours had acquired 5 sound changes (younger *Farningham 1880*) and 3½ (contemporary *Stoke 1868*).

The vowel diagram in Figure 4 looks superficially like an RP vowel diagram (Figure 6), but that is coincidental because both are variants of SBE.

For comparison, Figure 6 shows a vowel diagram for 19th century RP (Harold Macmillan, 1894-1986, UK prime minister 1957-63). He too has TRAP near DRESS, and an [ɔː]-like THOUGHT, but three of the SED informants had already shifted TRAP away from DRESS (higher F1) and five had already shifted THOUGHT to [oː]. RP would not shift TRAP until the early decades of the 20th century (Fabricius 2007). Macmillan also has similar timbre for STRUT and BATH, but darker than Wells (more [ɑ]-like with F2 around 1100Hz). Macmillan has [ɒ]-like RP LOT, similar to Wells' [ɑ] but lower F2. Finally, Macmillan has a darker [u]-like GOOSE than Wells with F2<1300Hz.

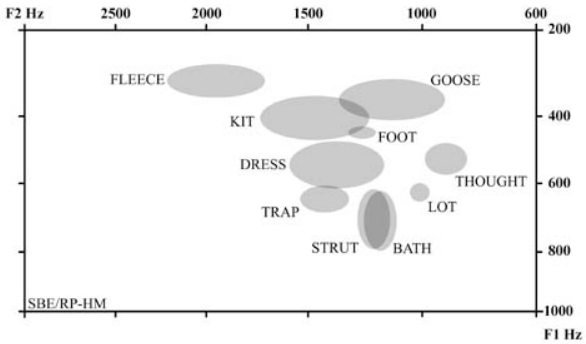

Figure 6. F1/F2 diagram for Harold Macmillan's 19th century RP monophthongs.

There are extracts from H. G. Wells speech recordings online at:
http://swphonetics.com/2015/02/10/halfway-to-estuary-english-h-g-wells/

References

Boersma, Paul & David Weeninck. 2014. *Praat: Doing phonetics by computer* (version 5.3), http://www.fon.hum.uva.nl/praat/, accessed September 2014.

British Library. Recordings of English accents. http://sounds.bl.uk/Accents-and-dialects. Accessed October 2014.

BBC Archive. 1931. Russia in the melting pot, radio broadcast by H. G. Wells. http://www.bbc.co.uk/archive/hg_wells/12401.shtml (accessed in January 2015).

Ellis, Alexander. 1889. *On Early English Pronunciation*. Teubner, Vol. 5.

Fabricius, Anne H. 2007. Variation and change in the TRAP and STRUT vowels of RP: a real time comparison of five acoustic data sets. *Journal of the International Phonetic Association* 37:293-320.

Gimson, A. C. 1962. *An Introduction to the Pronunciation of English*. London: Arnold.

Jones, Daniel. 1918. *An Outline of English Phonetics*. Marburg: Teubner. Then Heffer, Cambridge, to the 9th edn. (1964).

Masters, John W. 1820. *Dick and Sal at Canterbury Fair*. Canterbury. Reproduced by Parish & Shaw (1888).

Monty Python's Flying Circus. 1971. *Teach Yourself Heath*, bonus extra CD accompanying *Monty Python's Previous Record*. Track 41 in the CD reissued in 2014, http://www.montypythononlinestore.com/*/CD/Monty-Python-s-Previous-Record/3DC7040Z000. Accessed November 2014.

Mugglestone, Lynda. 2003. *Talking Proper*. Oxford: Oxford University Press.

Parish, William & William Shaw. 1888. *A Dictionary of the Kentish Dialect*. Lewes: Farncombe.

Robinson, Jonathon. -. The Survey of English Dialects. Leeds University Library. http://library.leeds.ac.uk/multimedia/imu/2210/SEDIM.pdf (accessed March 2015).

Rosewarne, David. 1984. Estuary English. *Times Educational Supplement* (19 October).

Sweet, Henry. 1892. *Primer of Phonetics*. Oxford: Clarendon.

Wales, Katie. 2006. *Northern English*. Cambridge UK: Cambridge University Press.

Wells J (1982). *Accents of English*. Cambridge UK: Cambridge University Press.

Wood, Sidney A. J. Forthcoming. How Estuary English spread through nineteenth century Kent.