The perceptual importance of falling pitch for speakers from different dialects of Swedish

Gosselke Berthelsen, Sabine

Published in:
Proceedings of Fonetik 2021

2021

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Gosselke Berthelsen, S. (2021). The perceptual importance of falling pitch for speakers from different dialects of Swedish. In M. Roll, & A. Hjortdal (Eds.), *Proceedings of Fonetik 2021: Lund, June 8–9, 2021* (Vol. 56, pp. 7-13). (Lund Working Papers in Linguistics; Vol. 56). Center for Language and Literature, Lund University. https://journals.lub.lu.se/LWPL/article/view/23266/20725

Total number of authors:
1

The perceptual importance of falling pitch for speakers from different dialects of Swedish

Sabine Gosselke Berthelsen
Centre for Languages and Literature, Lund University

Abstract

Falling pitch has long been argued to be a key feature in the distinction of Swedish pitch accents. In this paper, neurophysiological and behavioural evidence substantiating the perceptual importance of HL pitch contours at the word level is discussed. When presented with novel words, Swedes, regard-less of dialect, preattentively distinguished meaningful words with falling pitch. Falling pitch was also facilitative for tone mismatch detection. These responses to meaningful HL contours in foreign words were likely based on transfer from the native tongue, thus emphasising the importance of falling pitch in the discrimination of word accents in four Swedish dialect areas. However, while responses to falls were facilitated across dialects, additional dialect-dependent facilitation effects were found.

Introduction

Swedish has two pitch accents, accent 1 and accent 2, which are realised on word stems and affected by affixes (cf. Riad, 1998). Their pitch patterns differ systematically between dialects. Five dialect areas are distinguished in this respect: type 0, type 1A and 1B, and type 2A and 2B (cf. Gårding, 1973). Type 0 does not differentiate accent 1 and 2. The remaining dialect types are originally named with respect to the number of peaks that accent 2 receives in citation form (or focal position). A third word accent type is tentatively included for Gothenburg Swedish: type 3. Here, accent 1 and 2 have both been shown to elicit double peak contours in both focal and non-focal position (e.g., Segerup, 2005).

Word accents clearly differ within dialects, on the one hand, and between dialects, on the other. Yet, it is often argued that this variability is entirely based on timing differences and that all word accents are essentially different encodings of an underlying HL contour (e.g., Bruce, 1983, 2005; Bruce & Gårding, 1978). Thus, regardless of dialect, the onset of the word accent fall is earlier for accent 1 than for accent 2 (cf. *Figure 1*). Besides this crucial difference distinguishing the two word accents, there are further timing differences that distinguish the different dialect types. Thus, the onset of the word accent falls is earlier in type 2 dialects than in type 1 dialects. In type 2 dialects, the HL onset is in fact so early that accent 1 is realised as a low tone ([H]*L). The late onset in type 1 varieties, on the other hand, typically enforces the inclusion of a pitch rise from the previous word's L ([L]*HL). Regarding the A- and B-subtypes within type 1 and type 2, A has an earlier fall than B (Bruce, 1983, 2005). The double-peaked word accents in Gothenburg Swedish do not easily fit into the general word accent typology with respect to timing or pitch movement. Yet, there is a small but important difference in the timing of the fall in the stressed syllable between accent 1 and accent 2 even in this otherwise atypical dialect.

While the HL contour is likely the underlying feature driving word accent distinction within and between dialects overall, the vastly different realisations invite for speculations about the factual perceptual relevance of the fall and whether it is equally important in all dialects. It is conceivable, for example, that listeners from type 2 dialect areas perceive the difference between word accents as a high – low contrast or that speakers of type 1 dialects make use of the rising part of the contour to distinguish words, as the remnant rise precedes the H*L.

A reliable way of investigating the perceptual importance of pitch patterns in a population is to present speakers with a foreign tone system and study acquisition of and interaction with the new tones. Acquired perceptual strategies for the native language (i.e., specialised neural circuits) involuntarily guide the processing of foreign language input, at least initially. Swedish has recently received some attention in this context where it has been shown that Swedish speakers outperform speakers from non-tonal languages in discrimination or processing of foreign tones (e.g., Burnham et al., 2014; Gosselke Berthelsen, et al., 2020). Interestingly, this is the case when tones are embedded in a linguistic context but does not hold for musical tones.

More closely investigating not simply the advantage of having tones in the native language (L1) but looking at specific pitch patterns, it has further been illustrated that Swedish second language (L2) learners highly accurately identify falling L2 tones (i.e., Mandarin T4; Gao, 2016, 2019). Accuracies for L2 fall-rising tones (T3) are also high. It is not certain whether the latter is due to the presence of the initial fall or related to the fact that the fall-rise contour is reminiscent of the combined pattern for word accent and focus tone (in some varieties). Importantly, high and rising L2 tones (T1 & T2) are recognised with considerably lower accuracy.

This behavioural advantage for falling tones, however, seems to disappear with high intensity training (Gosselke Berthelsen et al., 2021). Under such circumstances, Swedish native speakers did not perform differently in the detection of mismatches related to falls compared to rises and high or low tones. In fact, their behavioural responses were statistically indistinguishable from those of non-tonal participants. Yet, while the behavioural measures were not indicative of a perceptual advantage for falls, a fall-facilitated neurophysiological response was observed just 50 milliseconds after the onset of the tone. This relatively newly discovered response has been associated with a lexical gating process, distinguishing words from nonwords and grammatical from ungrammatical forms (e.g., Herrmann et al., 2009; MacGregor et al., 2012; Shtyrov & Lenzen, 2017). A reduction in this EEG component, argued to be related to eased preconscious processing, was found for falls in meaningful words. This suggests that Swedish listeners are conditioned from their L1 to automatically use falling pitch as a cue to lexicality status. A process, that they preconsciously make use of even in the perception of foreign tones in a meaningful linguistic context.

Somewhat surprisingly maybe, none of the presented studies on Swedish native speakers' identification and processing of foreign tone offer clear information about the participants' dialectal background. Instead, they choose to treat Swedes as a uniform group with respect to word-level pitch experience. A choice that is presumably seen as justified by the suggestion that all word accent realisations are instantiations of HL contours. However, the differences in pitch realisations at the word level as a result of timing differences and interactions with sentence-level pitch (e.g., focus tones) might in fact result in differences in perceptual cues for word discrimination in the different dialects. The present paper presents a first attempt at bridging this gap and investigating if there are dialectal differences concerning perceptually important pitch patterns. For this purpose, behavioural and neurophysiological data from Gosselke Berthelsen et al. (2021) are analysed according to the participants' native dialect and emerging patterns for the lexical gating effect are discussed.

Methods

Participants

Twenty-three right-handed participants (aged 19-29, *M* = 23.7; 12 female) were informants in the present study. They were all native speakers of Swedish and came from three different dialect areas: South Swedish (type 1A), Central Swedish (type 2A, referred to by Bruce (e.g., 2005) as East Swedish), and West Swedish (type 2B). West Swedish was further separated into West Swedish and Gothenburg Swedish (type 3). Participants' dialects were determined in accordance with self-reported native dialect and primary area of residence, *Figure 1*. Note that participants moved within and across dialect borders (number of moves = 1-6; *M* = 2.8), including a move to the Lund area in the south of Sweden, where all participant were residing at the time of testing.

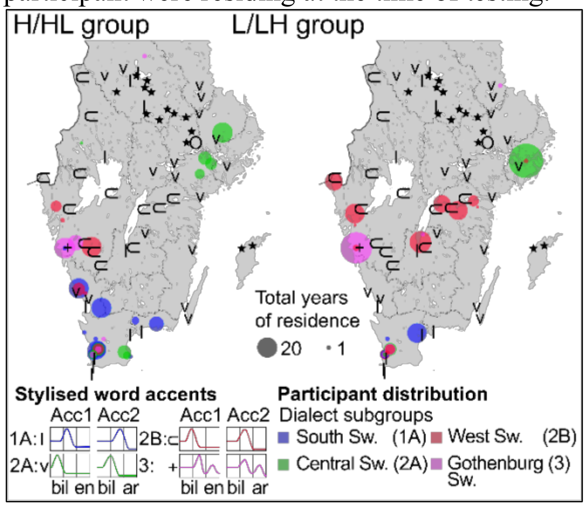

Figure 1. Maps for both groups indicating subgroups' residence distribution as well as Swedish word accent patterns. Word accent distribution adapted from Gårding & Lindblad (1973). Stylised, non-focal word accents modelled based on Bruce (1983) and Segerup (2005).

Participants were divided into two groups. They were tasked with learning novel foreign words with H* and H*L tones (H/HL group) or L* and L*H tones (L/LH group), respectively. Groups were matched for age ($M_{H/HL}$ = 23.4 years; $M_{L/LH}$ = 24.0 years) gender (H/HL: 6 female; L/LH: 6 female), and a large number of other background factors, such as working memory and socioeconomic status. Both learner groups could be divided into four subgroups representing the four dialectal areas, see *Table 1.*

Table 1. Number of participants per dialect.

Dialect	H/HL	L/LH
South Swedish	4	1
Central Swedish	3	3
West Swedish	2	5
Gothenburg Swedish	2	3

Materials

Twenty-four auditory, monosyllabic pseudowords were created for the experiment, *Table 2*. Although Swedish does not have differential pitch on monosyllables (they uniformly bear accent 1, e.g., bil_1, 'car'), monosyllables were deemed suitable for this investigation as the pitch accent on a monosyllabic stem is systematically altered when suffixes are added (e.g., bil_2*-ar*, 'cars', bil_2*-bälte*, 'seatbelt', but bil_1*-en*, 'the car', cf. Riad, 1998). In the present stimuli, rather than adding pitch-altering suffixes, the tones themselves served as grammatical suprafixes, see procedure below.

Table 2. Pseudowords used in the experiment.

i/u word pairs	a/ɛ word pairs
dif / duf	dap / dɛp
fif / fuf	fap / fɛp
kip / kup	kaf / kɛf
lir / lur	lap / lɛp
sis / sus	sap / sɛp
tip / tup	taf / tɛf

Half of the auditory stimuli were presented as novel words while the other half were presented as nonwords. A pseudoword learning paradigm was selected in order to guarantee equal exposure to the foreign words in which lexicality effects were to be studied. Word division was pseudorandom, such that i/u word pairs and a/ɛ word pairs were used as novel words or nonwords in an equal number of participants. In Praat (Boersma, 2001), four tones (H*, H*L, L*, L*H, see *Figure 2*) were added to the words in the following way: For any given participant, H*/H*L pitch would either be added to the novel words, assigning the participant to the H/HL group, or with nonwords, assigning the participant to the L/LH group. The L*/L*H would then be added to the other set of words: nonwords or novel words, respectively.

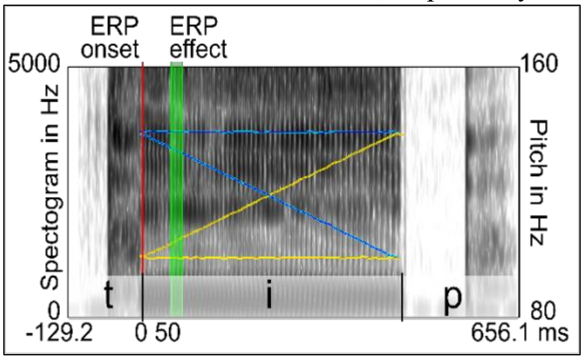

Figure 2. Spectrogram for example word /ti:p/ with the four pitch contours (H, H*L, L*, L*H).*

Procedure

Participants were presented with all words a total of 60 times, i.e. 30 times each on two consecutive days. All novel words were always followed by pictures of people in different professions to assign meaning to the words. The tones were associated with grammatical meaning, such that the difference between H* and H*L in the H/HL group or L* and L*H in the L/LH group would either be associated with grammatical gender or number distinctions (i.e., singular vs plural or feminine vs masculine). This was done to assimilate the function of pitch on Swedish words while keeping an easily controllable, monosyllabic word structure. Participants were told explicitly that the words that they would learn would contain gender and number markers. To uphold the participants' attention and to assess learning progress, in some trials (12.5%), the novel words were presented with a wrong picture. Participants were prompted to indicate incorrectly matched trials by a button press. Behavioural measures documenting the recognition of pictures that incorrectly matched the preceding word's tone will be discussed in the preliminary results.

Electroencephalography

During the experiment, EEG data was recorded from 64 Ag-AgCl EEG electrodes (EASYCAP GmbH, Herrsching, Germany) with a SynAmps^2 amplifier and Curry Neuroimaging Suite 7 software (Compumedics Neuroscan, Victoria, Australia). Two bipolar eye channels (EOG) were included. Left mastoid (M1) served as online reference and

AFz as ground. Sampling rate was 500 Hz and a low pass filter of 200 Hz was run online. Offline, the data was filtered with a 0.01 Hz high pass and a 30 Hz low pass filter. Epochs of 1200 ms were extracted at vowel/tone onset (cf. *Figure 2*) including a 200-ms baseline. Eye artefacts were corrected through independent component analysis (ICA, Jung et al., 2000). Epochs exceeding ±100 µV were excluded (M = 1.09%; SD = 0.72).

Preliminary analysis and results

To investigate the potential effect of dialect in the dataset, the behavioural measures response times and response accuracy were divided into group averages for the different dialects, see *Table 3* and *Table 4*. The H/HL group overall has faster response times (M = 1615 ms) and higher response accuracy (M = 79%) than the L/LH group (M_{RT} = 1889 ms; M_{ACC} = 55 %). Further, mismatches with the H*L pitch pattern (M = 1587 ms) are recognised more quickly than mismatches with any other tone type (M_{H*} = 1644 ms; M_{L*H} = 2018 ms; M_{L*} = 1760 ms). With respect to accuracy, there is virtually no differences between H*L tone mismatch recognition (M = 78.67%) and the detection of mismatches with H* tones (M = 78.68%). The two do, however, clearly differ from L*H mismatches (M = 56%) and L* mismatches M = 55%). Note that Gothenburg Swedish differs from the other dialects and does not favour H*L tones.

Table 3. Response accuracy for recognition of tone mismatches by dialect.

	H/HL group		L/LH group	
Dialect	H*L	H*	L*H	L*
South	64%	59%	34%	22%
Central	95%	94%	46%	49%
West	83%	88%	69%	65%
Gothenburg	73%	74%	75%	81%

Table 4. Response times in milliseconds for recognition of tone mismatches by dialect.

	H/HL group		L/LH group	
Dialect	H*L	H*	L*H	L*
South	1483	1359	1635	1503
Central	1213	1401	2654	2173
West	1295	1657	1741	1737
Gothenburg	2357	2167	2041	1626

Further, to study the participants' neurophysiological responses, effect amplitudes for the significant electrode cluster (FC2, FC4, C1, Cz, C2, CP1, CPz, and CP2) and time window (50-70 ms) from Gosselke Berthelsen et al., (2021) were separated into dialects, see *Table 3*. Mean amplitudes showed a reduced effect for H*L tones when they were part of meaningful new words (H/HL group) regardless of dialect (M = −0.76 µV). This reduction emerged in comparison to both nonwords (M_{L*H} = −1.04 µV; M_{L*} = −0.88 µV) and novel words with H* pitch (M = −0.86 µV). When presented as part of nonwords (L/LH group), the H*L tone (M = −1.06 µV) was not preattentively differentiated from the other tones (novel words: M_{L*H} = −0.98 µV; M_{L*} = −1.00 µV; nonwords: M_{H*} = −1.06 µV). The L/LH group instead showed a tendency towards a global amplitude reduction for newly learned real words (M = −0.99 µV) compared to nonwords (M = −1.06 µV). For Central Swedish speakers (type 2A), in addition to the reduced amplitude for the fall in novel words, there was also a noticeable amplitude reduction for L* and H* tones on novel words.

Table 5. Average amplitude in (µV) of all cluster electrodes in the 50-70 ms time-window for all dialect sub-groups by group and pitch pattern.

HL/H group	Words		Nonwords	
Dialect	**H*L**	H*	L*H	L*
South	**−1.05**	−1.19	−1.15	−1.16
Central	**−0.17**	**−0.19**	−0.82	−0.64
West	**−0.98**	−1.06	−1.25	−1.02
Gothenburg	**−0.83**	−1.00	−0.95	−0.71
LH/L group	**Words**		**Nonwords**	
Dialect	L*H	L*	**H*L**	H*
South	−0.81	−0.90	**−0.78**	−0.88
Central	−0.92	**−0.79**	**−1.09**	−1.07
West	−1.55	−1.57	**−1.70**	−1.66
Gothenburg	−0.65	−0.74	**−0.67**	−0.62

Discussion

The results presented here emphasise the suggested special status of the falling tone at word level for Swedish listeners. Swedes from most dialect areas could most easily identify incorrectly matched pictures when the mismatches were based on H*L tones. More directly measuring the pitch itself, neurophysiological results showed that H*L pitch, when presented on a meaningful word, reduced the early EEG component related to lexical gating. This reduction, arguably related to ease of processing, was found in listeners from all four dialect areas in the current study. This demonstrates the perceptual importance of pitch falls at the word level in Swedish regardless of dialect, strengthening the claim that it is indeed the fall that primarily distinguishes word accents regardless of their realisation in the complex interplay with timing and sentence level pitch (e.g., Bruce, 1983, 2005; Bruce & Gårding, 1978).

Given its close relation with word stems and affixes in Swedish, it is not surprising that the H*L tone could only serve as a cue for lexicality effects when it was added to meaningful words. When it had no meaning or function, it was too dissimilar from the native tones to affect L2 processing. Thus, in conditions where L* and L*H were associated with novel words, a tone-independent but substantially smaller effect of lexicality was observed. Listeners less successfully learned to use other differences in the speech input (i.e., L* and L*H tones or vowels) in absence of an H*L cue. This is substantiated by the behavioural data which shows reduced mismatch detection accuracy and prolonged response times for learners who had to rely on L* and L*H tones during word acquisition.

An intriguing pattern was observed for the Central Swedish subgroups. In addition to the amplitude reduction for falls, they also showed a reduced amplitude for high and low tones when these were part of meaningful words. Central Swedish (type 2A) is argued to have the earliest fall onset both for accent 1 and accent 2 (cf. Bruce, 1983). As such, accent 1 is realised as a low tone ([H]*L) in pre-focal position and for accent 2, the high tone is prominently associated with the stressed syllable (*HL). The preliminary results presented here indicate that the timing of the fall contour in Central Swedish has an impact on the perceptual prominence of pitch for native speakers of this dialect. To this effect, Central Swedish speakers preconsciously use the pitch fall as well as low and high level tones as perceptual cues for lexical distinction. Thus, response amplitudes for the early EEG component are reduced for all three pitch types, indicating eased processing due to transfer from the native dialect. This is partly corroborated in the behavioural responses. The Central Swedish participants in the H/HL group have by far the highest accuracy and fastest response times overall. This is likely due to the fact that they can preconsciously assess not only the H*L tone but also the H* tone. Further, those Central Swedish participants who learned words with L* and L*H tones have better response accuracy and response times for the L* tones than the L*H tones. However, overall accuracy is comparably low and response times are comparably long in this subgroup.

Interestingly, West Swedish (type 2B) did not share the Central Swedish (type 2A) emphasis on low and high tones. The relatively later onset of the word accent fall in this variety likely diminishes its potential interpretation as a low tone.

Another group of participants that very clearly stood out in the present study was the Gothenburg Swedish variety (here tentatively referred to as type 3). While the expected reduction for the H*L pattern in novel words was found in the neurophysiological data, the largest reductions were evident for level tones in non-words. It is unclear why the reduction would occur in the non-words rather than the words. This confound was not resolved but rather strengthened by the behavioural data. In both Gothenburg Swedish sub-groups, level pitch could more quickly and accurately be used for tone mismatch detection than the H*L and L*H contour tones. In fact, the H*L tone pattern was found to be the least useful cue for mismatch detection. The differences between conditions and between groups were, however, rather small and overall accuracy was high. It is likely that the similarity of the word accent patterns in the Gothenburg in both focal and non-focal position dialect reduces the perceptual importance of the falling pitch movement cue both for preconscious lexicality effects and for mismatch detection. Instead, listeners are likely attentive to other cues. It has previously been argued that vowel length plays an important role Gothenburg Swedish (e.g., Segerup, 2004). It is possible that the listeners in the present study interpreted the level tones as longer and could therefore effectively use them as cues for mismatch detection. However, this does not explain the reverse lexicality effect for level tones (i.e., a reduction for non-words rather than words). Further research on pitch perception in Gothenburg Swedish is thus certainly needed to understand the present data. The 1B dialect type was not included in the present investigation due to lack of participants. However, its relative similarity with type 1A word accents tentatively suggests that the fall would have the same status as it does in the remaining dialects. However, it is of course possible that there are additional perceptual cues. As the dialect area with the latest onset for the word accent falls, the type 1B variety has the longest rise leading up to the H of the HL contour. Hence, rises might potentially play a larger role in this dialect than in the other dialects.

Not mentioned explicitly in the introduction, the two previous studies on Swedish learners identification of L2 tones carried out by Gao (i.e., Gao 2016, 2019) find slightly different results with respect to the two very accurate patterns (fall vs fall-rise). In one study (2019) the fall is the most accurate pattern while in the other (2016), the fall-rise is slightly more accurate than the fall. In light of the present results, one possible explanation for this minor difference might be the participants' dialect background. It is possible, and indeed to some degree indicated in the paper, that the 2016 study had a large number of participants from type 2B varieties where, in focal position, accent 1 is realised as a fall-rise and accent 2 has a double peak pattern. This might lead to high fall-rise accuracy. The 2019 study, on the other hand, might have been conducted with a larger number of participants from type 1 varieties who are not familiar with double peak word accents from their native dialect, leading to the expected higher accuracy for the fall.

In sum, the current pilot study into the role of dialectal variation for the perceptual salience of pitch cues at the word level found that falling tones on meaningful words boosted preconscious lexicality effects in all the four tested dialectal varieties and possibly beyond. The neurophysiological results emphasise the key function that falling pitch plays in Swedish speakers' word perception overall. Yet, native dialect pitch timings and pitch prominence can affect perceptual salience and potentially sensitise listeners to additional pitch cues or different prosodic cues (such as vowel duration). Thus, while the fall protruded as the most important perceptual cue for speakers of all tested tonal dialects of Swedish, care should be taken when treating Swedes from different dialect backgrounds as a uniform group with respect to word-level pitch processing. More extensive data, including more participants, more dialects, and even a larger variety of tested pitch contours, is needed to corroborate the present preliminary findings and suggestions.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bruce, G. (1983). Accentuation and timing in Swedish. *Folia Linguistica, 17*(1-4), 221-238. doi:10.1515/flin.1983.17.1-4.221

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing.* doi:10.1093/acprof:oso/9780199249633.003.0015

Bruce, G., & Gårding, E. (1978). A prosodic typology for Swedish dialects. *Working Papers. (Phonetics Laboratory Lund University), 17.*

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., . . . Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 36*, 1459-1491. doi:10.1017/S0142716414000496

Gao, M. (2016). Perception of lexical tones by Swedish learners of Mandarin. *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, 33-40.

Gao, M. (2019). Perception of Tone 1-Tone 2 Contrast by Swedish Learners of Mandarin Chinese. *FONETIK 2019*, *27*, 7-12.

Gårding, E. (1973). The Scandinavian word accents. *Working Papers. (Phonetics Laboratory Lund University), 8*, 1-119.

Gårding, E., & Lindblad. (1973). Constancy and variation in Swedish word accent patterns. *Working Papers. (Phonetics Laboratory Lund University), 7*, 36-110.

Gosselke Berthelsen, S., Horne, M., Shtyrov, Y., & Roll, M. (2020). Different neural mechanisms for rapid acquisition of words with grammatical tone in learners from tonal and non-tonal backgrounds: ERP evidence. *Brain Research, 1729*(146614). doi:10.1016/j.brainres.2019.146614

Gosselke Berthelsen, S., Horne, M., Shtyrov, Y., & Roll, M. (2021). Native language experience narrowly shapes pre-attentive foreign tone processing and governs rapid memory trace build-up: An ERP study. *Manuscript submitted*.

Herrmann, B., Maess, B., Hasting, A. S., & Friederici, A. D. (2009). Localization of the syntactic mismatch negativity in the temporal cortex: An MEG study. *NeuroImage, 48*, 590-600. doi:10.1016/j.neuroimage.2009.06.082

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163-178. doi:10.1111/1469-8986.3720163

MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications, 3*(711), 1:7. doi:10.1038/ncomms1715

Riad, T. (1998). The origin of Scandinavian tone accents. *Diachronica, 15*(1), 63-98. doi:10.1075/dia.15.1.04ria

Segerup, M. (2005). The interaction of word accent and quantity in Gothenburg Swedish. *FONETIK 2005*, 13-16.

Segerup, M. (2004). Gothenburg Swedish word accents: a fine distinction. *FONETIK 2004*, 28-31.

Shtyrov, Y., & Lenzen, M. (2017). First-pass neocortical processing of spoken language takes only 30 msec: Electropysiological evidence. *Cognitive Neuroscience, 8*(1), 24-38. doi:10.1080/17588928.2016.1156663