Perception of L2 lexical stress in words degraded by a cochlear implant simulation

Everhardt, Marita K.; Sarampalis, Anastasios; Coler, Matt; Başkent, Deniz; Lowie, Wander

[Link to publication in University of Groningen/UMCG research database](#)

PERCEPTION OF L2 LEXICAL STRESS IN WORDS DEGRADED BY A COCHLEAR IMPLANT SIMULATION

Marita Everhardt[1,2], Anastasios Sarampalis[3,2], Matt Coler[4], Deniz Başkent[5,2], Wander Lowie[1,2]

[1]Center for Language and Cognition Groningen, University of Groningen, The Netherlands
[2]Research School of Behavioural and Cognitive Neurosciences, University of Groningen, The Netherlands
[3]Department of Experimental Psychology, University of Groningen, The Netherlands
[4]Campus Fryslân, University of Groningen, The Netherlands
[5]Department of Otorhinolaryngology, University Medical Center Groningen, The Netherlands
m.k.everhardt@rug.nl; a.sarampalis@rug.nl; m.coler@rug.nl; d.baskent@rug.nl; w.m.lowie@rug.nl

ABSTRACT

This study investigates how the perception of L2 lexical stress is affected by an acoustic simulation of cochlear implants (CIs). We explore whether Dutch L2 learners of English are influenced by f_0 differences or a vowel quality contrast when identifying lexical stress in L2 English words degraded by a CI simulation and whether listeners transfer cue-weighting strategies of the L1 into the L2. Results indicate that the identification of lexical stress based on a five-by-two matrix varying in f_0 and vowel quality was, as hypothesized, strongly compromised in the CI simulation, but that the lexical stress identification strategies for neither the unprocessed nor the CI-simulated stimuli differed between L1 Dutch and L2 English. This suggests that the lexical stress identification strategies of the L1 may have been transferred into the L2 and that the CI simulation of the present study affected L1 Dutch and L2 English lexical stress perception similarly.

Keywords: perception; second language learners; prosody; lexical stress; cochlear implant simulation

1. INTRODUCTION

Cochlear implants (CIs; auditory prostheses which use electrodes to stimulate the auditory nerve directly) can restore hearing in deaf individuals, yet speech perception with electric hearing is less accurate than with normal acoustic hearing due to degradations in fine spectrotemporal detail [3]. In this study, we explore how a degradation by a CI simulation may affect the perception of lexical stress in English as a second language (L2) within a normal-hearing (NH) population. Lexical stress, where a given syllable of a word is identified as stressed depending on its perceived prominence relative to another syllable, can be signalled through prosodic cues such as variation in fundamental frequency (f_0), intensity, or duration, as well as through variation in vowel quality. In this study, we focus on two of the aforementioned cues that CI users perceive less accurately than NH listeners: f_0 and vowel quality [3, 10, 16], where stressed syllables are higher in average f_0 than unstressed syllables and show a (high) f_0 peak, whereas vowels in unstressed syllables are more centralised compared to their stressed counterpart [14]; the vowel is usually reduced to schwa [7].

Languages have been found to differ in the relative weight they attach to these cues: in Dutch, f_0 differences have a greater functional weight than the vowel quality contrast in signalling lexical stress, whereas in English the opposite is true [5, 14]. Moreover, the cue-weighting theory not only states that languages differ in the relative weight they attach to cues, but also that listeners may transfer the cue-weighting strategies from their native language (L1) into their L2 [9, 12]. Accordingly, Dutch L2 learners of English are expected to rely more on f_0 differences when identifying lexical stress in L2 English than on the vowel quality contrast.

The question in this paper is how the transfer of cue-weighting strategies may be affected by degradations imposed by a CI simulation. Research has indicated that CI users can more accurately differentiate between vowels than f_0-related prosodic contrasts both in isolation [10] and in a single-word context [16], suggesting there is a perceptual hierarchy between these cues where the vowel quality contrast is more perceptually salient in electric hearing than f_0 differences. While CI users adapt to degradations and develop long-term perceptual strategies to compensate for them [2], a CI simulation is a good first step to identify approximate acute effects of acoustic-phonetic degradations of electric hearing. As a result, if listeners mostly rely on the cue that is perceptually more salient, this would mean that Dutch L2 learners of English mainly rely on the vowel quality contrast when identifying lexical stress in L2 English CI-simulated words, which is the opposite of what the cue-weighting theory predicts.

In this study, we investigate whether the perception of L2 English lexical stress in CI-simulated words is mostly influenced by a transfer of cue-weighting strategies or by the perceptual salience of available cues to lexical stress, where we consider

how the lexical stress identification patterns for CI-simulated words compares to these patterns for non-CI-simulated (hereafter: unprocessed) words and how the identification patterns of Dutch L2 learners of English listening to L2 English words compares to these patterns for L1 Dutch words. We hypothesized that identification of L2 English lexical stress in CI-simulated words would be strongly compromised due to degradations in fine spectrotemporal detail [3, 10, 16] and that Dutch L2 learners of English – despite the perceptual salience of the vowel quality contrast observed in CI users [10, 16] – would mainly rely on f_0 differences and ignore the vowel quality contrast due to a transfer of cue-weighting strategies [5, 9, 12, 14], thus outweighing the perceptual salience of available cues.

2. METHOD

2.1. Participants

Twenty-one first-year secondary school students (5 male, 16 female) aged between 12.0 and 13.1 years (*M*: 12.5, *SD*: 0.28) participated in the study, who are representative of late L2 learners whose L1 is (largely) established at the onset of L2 acquisition. Inclusion criteria were: being an L1 speaker of Dutch, learning L2 English at school (with no more than three years of formal instruction; mean age at onset L2 instruction: 10.2, *SD*: 1.64), and having NH (pure-tone thresholds better than 20 dB HL at audiometric frequencies between 250 and 8000 Hz).

2.2. Stimuli

The English stimuli comprised of disyllabic lexical stress pairs, where the stress falls either on the first syllable (strong-weak; SW) or on the second syllable (weak-strong; WS), such as <u>contract</u> vs. <u>contract</u> (underlined syllables are stressed). The Dutch stimuli comprised of both disyllabic (e.g. <u>misbruik</u> vs. <u>misbruik</u>) and trisyllabic lexical stress pairs. Similar to the disyllabic pairs, the stress in trisyllabic pairs falls either on the first syllable (strong-weak-weak; SWW) or on the second syllable (weak-strong-weak; WSW), such as <u>voorkomen</u> vs. <u>voorkomen</u>.

Five adult female L1 speakers of English and five adult female L1 speakers of Dutch recorded ten English and ten Dutch lexical stress pairs respectively (48-kHz sampling frequency, 16-bit). The ten pairs per language were evenly divided amongst the L1 speakers, such that each speaker would contribute two lexical stress pairs to the stimuli set. The forty source recordings (2 languages x 5 speakers x 2 words x 2 stress patterns) were subsequently processed in three steps, creating CI-simulated five-by-two matrices varying in f_0 and vowel quality.

Firstly, we normalised duration and intensity in Praat [4], as to disregard durational and intensity cues to lexical stress. We normalised the syllable duration of each stimulus such that the duration of each syllable corresponded to the average duration of that syllable across stress patterns. Calculations of average syllable duration were based on acoustic measurements of the selected source recordings. Similarly, we normalised the syllable intensity of each stimulus such that the intensity of each syllable corresponded to the average (root-mean-square; RMS) intensity of that syllable across stress patterns. Calculations of average syllable intensity were based on acoustic measurements of mean syllable intensity, measured after normalisation of syllable duration.

Secondly, we created five-step f_0 continua, where one end of a continuum represented the SW f_0 pattern and the other end the WS f_0 pattern (SW and WS henceforth refer to both SW–WS and SWW–WSW). Calculations of the five-step f_0 continua were based on acoustic measurements of f_0 in semitones (reference level: 100 Hz) at the start and the end of the voiced part of each syllable, measured after normalisation of duration and intensity. The five-step f_0-continuum manipulations were carried out using PSOLA implemented in Praat [4] and were applied to both the SW and WS source recordings. As these source recordings – after normalisation of duration and intensity – varied only in f_0 and vowel quality, the application of the f_0 manipulations to both sources created a five-by-two matrix of each stimulus.

Lastly, we created an acoustic CI simulation of each stimulus by means of a vocoder [8] implemented in MATLAB [13]. Vocoder parameters were based on a previous vocoder study where f_0 cues were manipulated [6]. We modified the parameters to further limit spectral resolution and temporal envelope cues by reducing number of channels and the envelope filter cut-off respectively as to minimize acoustic cues related to f_0 and vowel quality. Specifically, we used a 6-channel noise-band vocoder with a bandwidth of 250-8700 Hz and Greenwood map, using zero-phase 12th order Butterworth filters with matching analysis and synthesis filters. The temporal envelope was extracted by half-wave rectification and low-pass filtering at a cut-off of 100 Hz using a zero-phase 4th order Butterworth filter.

2.3. Procedure

Participants completed an identification task in which they listened to unprocessed and vocoded stimuli in L1 Dutch and L2 English after which they had to indicate for each stimulus whether the first or the second syllable was stressed. Participants were informed that they would hear same-word stimuli

more than once, but that this did not necessarily mean that the stress pattern was the same. The stimuli were presented over headphones at a comfortable hearing level and responses were given by pressing a key on a keyboard. Response choices were automatically recorded in OpenSesame [11].

The experiment was divided into four blocks, one for each processing type per language. The block order was pseudo-randomised such that the order of language was counterbalanced between participants and the order of processing within each language was subsequently counterbalanced. Stimuli within a block were presented in randomised order, where immediate succession of same-word stimuli was prevented. Each block started with a practice session, where participants were introduced to the extreme points of the five-by-two matrices. In the practice sessions, participants received feedback. In the experiment proper, no feedback was given.
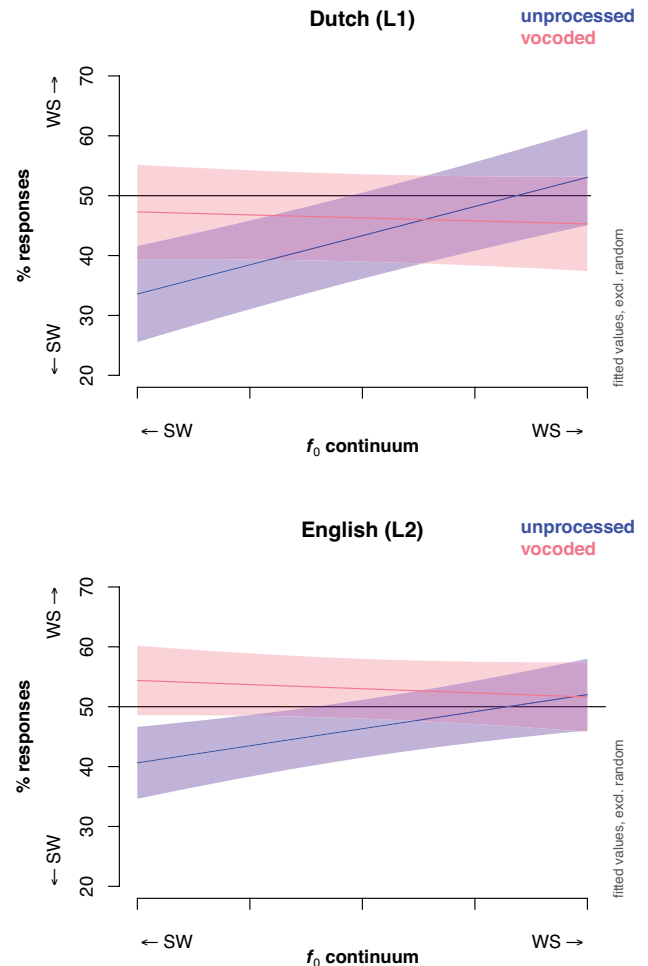
3. RESULTS

3.1. Identification patterns

To assess if Dutch L2 learners of English made use of f_0 differences when identifying lexical stress, we fitted a generalized additive model (GAM) in R using the *mgcv* package [17], with a smooth over the f_0 continuum as a predictor for response and separately for unprocessed and vocoded stimuli per language. The model was collapsed over vowel quality, as model comparison – using the Akaike Information Criterion (AIC) [1] – revealed that the simpler model without the vowel quality contrast was preferred (AIC difference: -7.08), indicating that the identification patterns for SW-vowel stimuli did not significantly differ from the patterns for WS-vowel stimuli. The model also included a by-participant factor smooth for the interaction between processing and vowel, as well as a by-word factor smooth for language.

The identification patterns over the f_0 continuum for unprocessed and vocoded stimuli per language are presented in Figure 1 (visualized using the *itsadug* package [15]). The model described above revealed a significant smooth effect over the f_0 continuum for the unprocessed stimuli in both Dutch (F = 28.95, p < .001) and English (F = 9.90, p = .002), indicating that Dutch L2 learners of English made use of f_0 differences when identifying lexical stress in both languages, where later f_0 continuum points generated more WS responses compared to earlier points. The smooth effect over the f_0 continuum for the vocoded stimuli did not reach significance for either Dutch or English (p > .05), indicating that Dutch L2 learners of English were unable to make use of f_0 differences when identifying lexical stress in vocoded words.
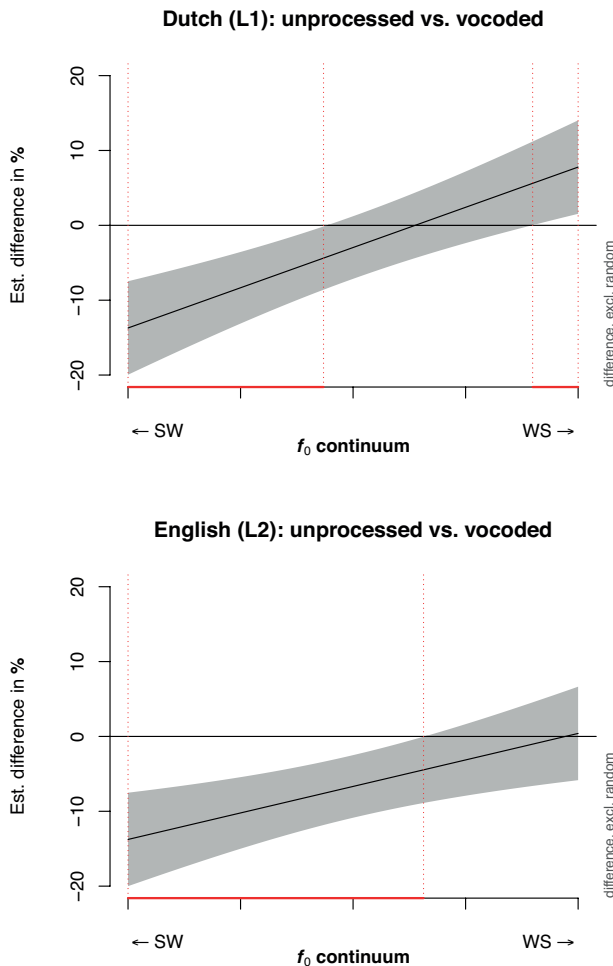
3.2. Processing contrast

To assess if the identification patterns for unprocessed and vocoded stimuli (as outlined in 3.1 and in Figure 1) significantly differed, we fitted a model with a difference smooth between unprocessed and vocoded stimuli per language (presented in Figure 2). The model revealed a significant difference between these identification patterns for both Dutch (F = 11.29, p < .001) and English (F = 9.38, p < .001), indicating that the CI simulation of the present study significantly influenced the lexical stress identification patterns in both languages. The difference in identification patterns between unprocessed and vocoded words mostly lies in the earlier f_0 continuum points, as can be seen in both Figure 1 and Figure 2; earlier f_0 continuum points generated significantly more SW responses in unprocessed words compared to vocoded words.

Figure 2: Estimated difference (incl. 95% confidence bands) in percentage of lexical stress identification responses over the five-step f_0 continuum (left: SW, right: WS) comparing unprocessed with vocoded speech for L1 Dutch and L2 English stimuli (red line on x-axis and vertical dotted lines indicate significant differences).

Dutch (L1): unprocessed vs. vocoded

English (L2): unprocessed vs. vocoded

3.3. Language difference

To assess if the identification patterns for unprocessed and vocoded stimuli differed between L1 Dutch and L2 English and thereby assess how the CI simulation affected the transfer of cue-weighting strategies, we fitted models with difference smooths between Dutch and English per processing type, as well as for the processing contrast. The difference between identification patterns of L1 Dutch and L2 English for neither unprocessed nor vocoded words reached significance ($p > .05$), nor did the language difference for the processing contrast ($p > .05$), indicating that Dutch L2 learners of English transferred identification strategies from L1 Dutch to L2 English and that the CI simulation of the present study did not influence this transfer.

4. DISCUSSION

In this study, we explored how degradations by a CI simulation may acutely influence L2 lexical stress identification patterns. Results showed that Dutch L2 learners of English made use of f_0 differences – the cue with the greatest functional weight in Dutch [5, 14] – in both L1 Dutch and L2 English when identifying lexical stress in unprocessed words. Yet, they were unable to make use of this cue in CI-simulated words; responses for CI-simulated stimuli were around chance (50%) for all f_0 continuum points (see Figure 1). Further analyses revealed that the response patterns for both L1 Dutch and L2 English significantly differed between unprocessed and CI-simulated words; the identification of lexical stress based on f_0 differences was strongly compromised by the CI simulation. While there is a possibility that this finding is unique to the specific vocoder parameters of this study, it is consistent with previous research that suggests that the perception of f_0 differences is less accurate in electric hearing [3, 10, 16].

The significant influence of f_0 differences in both L1 Dutch and L2 English for unprocessed words indicates that the Dutch L2 learners of English made use of the cue with the greatest functional weight in the L1 in both their L1 and L2, suggesting that they may have transferred the L1 cue-weighting strategies into the L2 [5, 9, 12, 14]. Moreover, results showed that listeners did not make use of the vowel quality contrast – the cue with the greatest functional weight in English [5, 14] – when identifying L2 English lexical stress. This also implies that they did not conform to the cue-weighting strategies of the L2 in either unprocessed or CI-simulated speech, even though literature implies that the vowel quality contrast could be more perceptually salient than f_0 differences in electric hearing [10, 16]. That said, we acknowledge that a follow-up study with L1 English listeners is needed to further investigate this finding.

The lack of a significant difference in lexical stress identification patterns between L1 Dutch and L2 English in either unprocessed or CI-simulated speech (or the contrast) furthermore suggests that L1 Dutch identification strategies were also applied in L2 English, implying that the L1 cue-weighting strategies may have been transferred into the L2 without any influence from the CI simulation. We thus conclude from these preliminary data that the transfer of cue-weighting strategies could outweigh the perceptual salience of available cues to lexical stress in CI simulations, yet we acknowledge that different vocoder parameters might lead to different results. That said, the CI simulation of the present study affected L1 Dutch and L2 English lexical stress perception similarly.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.

[2] Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A., Gaudrain, E. 2016. Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved? *Trends in Hearing* 20, 1–16.

[3] Başkent, D., Gaudrain, E., Tamati, T. N., Wagner, A. 2016. Perception and psychoacoustics of speech in cochlear implant users. In: Cacace, A. T., de Kleine, E., Holt, A. G., van Dijk, P. (eds), *Scientific foundations of audiology: Perspectives from physics, biology, modelling, and medicine.* San Diego, CA: Plural Publishing Inc., 285–319.

[4] Boersma, P., Weenink, D. 2018. *Praat: Doing phonetics by computer (version 6.0.37).* Retrieved from http://www.praat.org/

[5] Cutler, A. 2009. Greater sensitivity to prosodic goodness in non-native than in native listeners. *Journal of the Acoustical Society of America* 125(6), 3522–3525.

[6] El Boghdady, N., Başkent, D., Gaudrain, E. 2018. Effect of frequency mismatch and band partitioning on vocal tract length perception in vocoder simulation of cochlear implant processing. *Journal of the Acoustical Society of America* 143(6), 3505–3519.

[7] Fear, B. D., Cutler, A., Butterfield, S. 1995. The strong-weak syllable distinction in English. *Journal of the Acoustical Society of America* 97(3), 1893–1904.

[8] Gaudrain, E. 2016. *Vocoder, v1.0.* Retrieved from https://github.com/egaudrain/vocoder

[9] Holt, L. L., Lotto, A. J. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America* 119(5), 3059–3071.

[10] Luo, X. Fu, Q.-J., Wu, H.-P., Hsu, C.-J. 2009. Concurrent-vowel and tone recognition in Mandarin-speaking cochlear implant users. *Hearing Research* 256(1-2), 75–84.

[11] Mathôt, S., Schreij, D., Theeuwes, J. 2012. OpenSesame: An open-source graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2), 314–324.

[12] Qin, Z., Chien, Y.-F., Tremblay, A. 2017. Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics* 38(3), 541–570.

[13] The Mathworks. 2018. *MATLAB and Statistics Toolbox (Release 2018a).* Retrieved from https://nl.mathworks.com

[14] Tremblay, A., Broersma, M., Coughlin, C. E. 2018. The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition* 21(3), 640–652.

[15] van Rij, J. C., Wieling, M. B., Baayen, R. H., van Rijn, D. H. 2017. *itsadug: Interpreting time series and autocorrelated data using GAMMs (version 2.3).* Retrieved from https://CRAN.R-project.org/package=itsadug

[16] van Zyl, M., Hanekom, J. J. 2013. Perception of vowels and prosody by cochlear implant recipients in noise. *Journal of Communication Disorders* 46(5-6), 449–464.

[17] Wood, S. N. 2011. *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation (version 1.8.26).* Retrieved from https://CRAN.R-project.org/package=mgcv