Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Interpretation of prosodically marked focus in cochlear implant-simulated speech by non-native listeners

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

[1]Center for Language and Cognition Groningen, University of Groningen, Netherlands
[2]Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands
[3]Research School of Behavioural and Cognitive Neurosciences, University of Groningen, Netherlands
[4]Department of Psychology, University of Groningen, Netherlands
[5]Campus Fryslân, University of Groningen, 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 assesses how a cochlear implant (CI) simulation influences the interpretation of prosodically marked linguistic focus in a non-native language. In an online experiment, two groups of normal-hearing native Dutch learners of English of different ages (12–14 year-old adolescents vs. 18+ year-old adults) and with different proficiency levels in English (A2 vs. B2/C1) were asked to listen to CI-simulated and non-CI-simulated English sentences differing in prosodically marked focus and indicate which of four possible context questions the speaker answered. Results show that, as expected, focus interpretation is significantly less accurate in the CI-simulated condition compared to the non-CI-simulated condition and that more proficient non-native listeners outperform less proficient non-native listeners. However, there was no interaction between the influence of the spectro-temporal degradation of the CI-simulated speech signal and that of the English proficiency level of the non-native listeners, suggesting that less proficient non-native listeners are not more strongly affected by the spectro-temporal degradation of an electric speech signal than more proficient non-native listeners.

Index Terms: perception, non-native listeners, linguistic focus, prosody, cochlear implant simulation

1. Introduction

Cochlear implants (CIs) are auditory prostheses that can partially restore hearing in individuals with sensorineural hearing loss. The electrode array inserted in the cochlea sends electrical speech cue signals to the auditory nerve directly. This electric input makes speech perception possible. However, electric hearing differs from normal acoustic hearing. An electric speech signal is highly lacking in fine spectro-temporal detail, resulting in a lower quality of transmission of acoustic cues, particularly fundamental frequency (f_0) [1, 2, 3]. Limited access to the prosody-related acoustic cue f_0 in an electric speech signal has been shown to compromise prosody identification in the native language [4]. The present study explores how a CI simulation (i.e., an acoustic approximation of electric hearing implemented by a noise-band vocoder) influences the interpretation of prosodic patterns—specifically the interpretation of prosodically marked linguistic focus—in a non-native language.

Linguistic focus highlights those words in speech that are new or otherwise important [5, 6, 7]. In English, linguistic focus is prosodically marked by the nuclear accent. That is, focused words are realised with localised phonetic prominence relative to the words around it, generally indicated by higher f_0, greater f_0 movement, increased intensity, and increased duration. By default, the word at the end of an intonation phrase carries the nuclear accent, unless a preceding word is in focus, in which case the focused word carries the nuclear accent [8, 6, 9]. The context of the discourse influences what part of an utterance is in focus [10, 11]. Consider the following example (note: $[...]_F$ indicates the focus and capitalisation indicates the nuclear accent):

1. John kicked the ball.
   (a) Who kicked the ball?
   $[\text{JOHN}]_F$ kicked the ball. (subject focus)
   (b) What did John do with the ball?
   John $[\text{KICKED}]_F$ the ball. (verb focus)
   (c) What did John kick?
   John kicked $[\text{the BALL}]_F$. (object focus)
   (d) What happened?
   $[\text{John kicked the BALL}]_F$. (broad focus)

The focus patterns in 1a–1d are in response to the preceding context questions. The subject focus (SF), verb focus (VF), and object focus (OF) responses are examples of narrow focus as only a single word is focused. No specific word is in focus for the broad focus (BF) response. As a result, the nuclear accent is by default on the phrase-final word, in this case the object.

Native listeners rapidly and efficiently recognise focused words and the discourse implications of the prosodically marked focus [12]. A recent study has, for instance, shown that native listeners can correctly interpret the position of focused words in spoken utterances with respect to context questions [10]. In the experiment, adult native (New Zealand) English listeners were asked to listen to canonical English sentences with the nuclear accent either on the subject or on the object and indicate which of two context questions was the most likely question given the way the speaker responds. These context questions were designed to either prompt the SF response or

10.21437/SpeechProsody.2022-15

the OF response. The results showed that listeners were more likely to select the SF question when the nuclear accent was on the subject and more likely to select the OF question when the nuclear accent was on the object, suggesting they correctly interpreted the focus position and were able to identify the correct context question.

Non-native listeners, however, recognise prosodic patterns less accurately and less efficiently [13, 14]. One study specifically showed that non-native listeners performed significantly less accurately compared to native listeners during a task where they had to indicate whether the prosodically marked focus pattern of a response was appropriate given a context question for matched and mismatched question-answer pairs in English. That is, whereas the adult native (American) English listeners were very likely to accept matched pairs and reject mismatched pairs, the Korean and Mandarin listeners had greater difficulty recognising the prosodically marked focus, mainly evidences by the lower accuracy in rejecting mismatched question-answer pairs [15]. Importantly, the study also showed that the greater the English proficiency of the Korean and Mandarin listeners, the higher the accuracy. This is in line with research in the segmental domain showing a beneficial effect of amount of experience with the non-native language [16]. Non-native prosodic focus interpretation thus appears to improve with increasing experience and proficiency in the non-native language.

Little is known about how prosodic focus interpretation for non-native language learners with different proficiency levels may be affected by the spectro-temporal degradation of an electric speech signal. For native listeners, studies have demonstrated that typical CI users (i.e., postlingually deafened and implanted adults) are less accurate in correctly identifying prosodically focused words compared to their normal-hearing (NH) peers [17, 18]. Comparable results were found for early deafened and implanted children [19], for prelingually deafened adolescent CI users [20], for adults who were deafened either prelingually or postlingually, but who were all implanted late [21], and for NH adult listeners listening to CI-simulated speech [22]. For non-native listeners, the influence of electric hearing on prosodic focus interpretation remains unclear. While there are no studies directly investigating this effect, speech-in-noise perception studies indicate that non-native listeners are more strongly affected by adverse listening conditions than native listeners [23], suggesting that non-native listeners would also be more strongly affected by spectro-temporal degradations.

The present study assesses in an online experiment how spectro-temporal degradations, similar to those that can occur in electric hearing, influence the interpretation of prosodically marked focus in a non-native language for non-native language learners with different proficiency levels (based on the amount of experience with non-native language learning in a school setting). Using a task that tests whether listeners can link the focus pattern to the correct context question [10], we investigate how accurately less proficient adolescent and more proficient adult native Dutch learners of English interpret prosodically marked focus in CI-simulated English sentences compared to how accurately these listeners interpret focus in non-CI-simulated English sentences. In line with previous research, we predict that the non-native listeners interpret prosodic focus less accurately in the CI-simulated sentences than in the non-CI-simulated sentences [22] and that the more proficient adult native Dutch learners of English will interpret prosodic focus in English sentences more accurately than the less proficient adolescent native Dutch learners of English [15].

2. Method

2.1. Participants

Two groups of participants took part in the online study: beginner native Dutch learners of English and more experienced and more proficient native Dutch learners of English. Eighteen 12–14 year-old (*M*: 13.5, *SD*: 0.93) adolescent secondary school students formed the group of beginner non-native language learners. They have a limited amount of experience with English learning in a school setting and have an estimated average Common European Framework of Reference for Languages (CEFR) [24] level of English listening skills of A2 [25]. Thirty-two 18+ year-old (*M*: 20.6, *SD*: 2.24; range: 18–30) adult first-year psychology students from the University of Groningen formed the group of the more proficient non-native language learners. They have more years of experience with English learning in a school setting and have an estimated CEFR level of at least B2/C1 as they have to meet these English language requirements of pre-university education in order to enter Dutch universities. To confirm normal hearing, participants were asked to successfully complete the Dutch version of the online digits-in-noise (DIN) test [26] before participation in the online study. All participants declared to meet this requirement. Written informed consent was obtained from all participants and from the parents/guardians of the adolescents before participation in the study. The adolescents were paid for their time and the adults received course credit.

2.2. Stimuli

Twenty-six English sentences were constructed (24 trial items, 2 practice items) with a subject-verb-article-object structure such as “John kicked the ball”. Four context questions were created for each sentence. These questions were designed to elicit a SF, VF, OF, or BF response and followed the pattern of the questions outlined in 1a–1d. The stimuli were recorded by two female and two male adult native speakers of (British) English. Each speaker recorded the 24 trial items and 2 practice items in all focus forms. To elicit natural focus responses, speakers responded to the context questions. The stimuli were recorded using a TASCAM DR-100 portable digital recorder with a Sennheiser 3865 condenser microphone at a sampling frequency of 48 kHz and sampling depth of 16 bit. The 96 trial stimuli (24 sentences x 4 focus types) were evenly divided between the speakers, such that each speaker contributed one focus type per sentence and a total of six SF, six VF, six OF, and six BF stimuli to the final set. Acoustic analyses of the stimuli will be presented in a forthcoming paper.

Acoustic CI simulations of the selected stimuli were created by means of a vocoder (version 1.0) [27] implemented in MATLAB (R2018a). Vocoded stimuli were created using an 8-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 160 Hz using a zero-phase 4th order Butterworth filter. These parameters resembled those of a previous study [28], but were modified to more closely approximate the perceptual abilities of the average CI listener by selecting an 8-channel vocoder and a cut-off frequency of 160 Hz [2, 29]. A pilot study confirmed that participants were able to perform the tasks described below in the vocoded condition and that the selected parameters would lead to accuracy scores above chance-level.

2.3. Procedure

Participants completed an online experiment assessing the influence of the CI simulation on the interpretation of prosodically marked focus in English sentences using a single-task and dual-task paradigm (note: the set-up and results of the dual-task paradigm assessing listening effort by the addition of a free recall task will be discussed in a forthcoming paper as they are out of the scope of this paper). The experiment was coded in jsPsych (version 6.1.0) [30] and data collection was managed through a JATOS server (version 3.5.5) [31].

The primary task of the experiment was a single-interval four-alternative forced-choice (1I-4AFC) focus interpretation task with unprocessed (non-vocoded) and vocoded stimuli. During each trial, participants were presented with an auditory stimulus after which they were asked to indicate which of four possible context questions the speaker answered. The four context questions were presented as stacked response buttons and were accompanied by the prompt "Which question did the speaker answer?" (Dutch: "Welke vraag heeft de spreker beantwoord?"). The stimuli were presented in randomised order with the constraint that immediate succession of same-sentence stimuli, regardless of the focus pattern, would not be possible.

The experiment was divided into four blocks, one for each processing condition per task paradigm. The block order was pseudo-randomised and participants were randomly assigned to one of the four possible block orders. Each block started with a practice session during which participants received feedback on the 1I-4AFC focus interpretation task. In the experiment proper, no feedback was given.

Participants were instructed to complete the online experiment in a quiet environment and were asked to use good quality headphones. Sound levels could be calibrated at the start of the experiment; participants were instructed to adjust the volume of their headphones or computer until they could hear a sample sentence at a clear and comfortable level and to not adjust the volume thereafter. Participants were also presented with a vocoded speech sample at the start of the experiment so they could familiarise themselves with vocoded speech.

3. Results

The 1I-4AFC task response data were analysed by fitting generalised linear mixed-effect models (GLMMs) in the R environment (version 4.1.2) using the *glmer* function of the *lme4* package (version 1.1-27.1) [32]. The prosodic focus interpretation patterns were fitted as the estimated probability of a correct response. The need for predictor variables and by-participant and by-stimulus random slopes was assessed through step-wise model comparisons using the *anova* function, starting from a basic model with only by-participant and by-stimulus random intercepts. The final model included an interaction between *processing* and *focus*, an interaction between *group* and *focus*, and an interaction between *paradigm* and *focus* (not discussed here), as well as by-participant and by-word random slopes for *processing* and *focus*. Post-hoc analyses were performed and visualised using the *emmeans* package (version 1.7.0) [33].

The model predictions in Figure 1 show that the probability of a correct response is lower in the vocoded condition than in the unprocessed condition (panel A) and that the more proficient adults outperform the less proficient adolescents (panel B). Post-hoc pairwise comparisons confirmed that the response probability is significantly lower for vocoded (vs. unprocessed) stimuli for all focus types (SF: $z = -9.06$, $p < .001$; VF: $z = -10.57$, $p < .001$; OF: $z = -5.35$, $p < .001$; BF: $z = -7.16$, $p < .001$). Moreover, the response probability is significantly higher for the more proficient adult (vs. less proficient adolescent) listeners for SF ($z = 4.88$, $p < .001$) and OF ($z = 4.03$, $p < .001$), but not for VF or BF ($p > .05$).

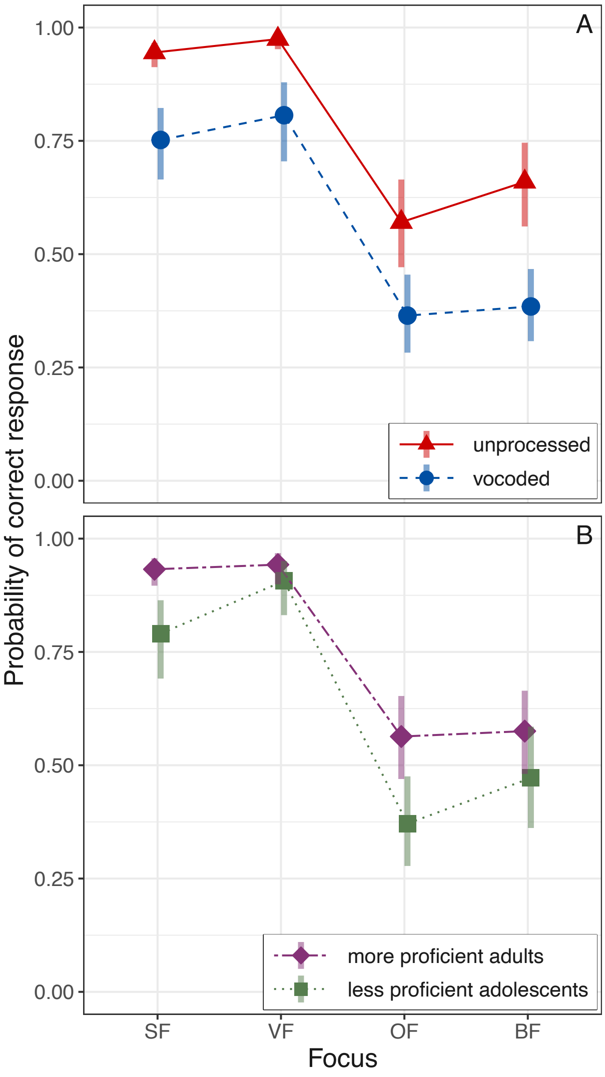

Figure 1: *Estimated probability of a correct response per focus type (SF, VF, OF, BF) for unprocessed (red triangles) and vocoded (blue circles) stimuli averaged across participant groups and task paradigms (panel A) and for more proficient adult (purple diamonds) and less proficient adolescent (green squares) non-native listeners averaged across processing conditions and task paradigms (panel B). Shaded areas show the 95% confidence intervals. BF: broad focus; OF: object focus; SF: subject focus; VF: verb focus.*

Figure 1 also shows that the probability of a correct response is relatively high for SF and VF, but much lower for OF and BF. This pattern is evident in both processing conditions and for both participant groups. Post-hoc analyses confirmed that the response probability is significantly lower for OF than for SF (unprocessed: $z = -9.07$, $p < .001$; vocoded: $z = -6.35$, $p < .001$; adults: $z = -8.57$, $p < .001$;

adolescents: $z = -6.34$, $p < .001$) or VF (unprocessed: $z = -9.13, p < .001$; vocoded: $z = -5.80, p < .001$; adults: $z = -7.07$, $p < .001$; adolescents: $z = -7.13$, $p < .001$) and also significantly lower for BF than for SF (unprocessed: $z = -7.33$, $p < .001$; vocoded: $z = -5.67$, $p < .001$; adults: $z = -7.67$, $p < .001$; adolescents: $z = -4.22$, $p < .001$) or VF (unprocessed: $z = -8.55$, $p < .001$; vocoded: $z = -5.92, p < .001$; adults: $z = -7.36, p < .001$; adolescents: $z = -6.45, p < .001$).

Regardless of the processing condition, the response probability for SF does not significantly differ from the probability for VF, nor does the probability for OF differ from the probability for BF ($p > .05$). Similarly, for the more proficient adults, the response probability for SF does not differ from the probability for VF and the probability for OF does not differ from the probability for BF ($p > .05$). For the less proficient adolescents, however, the response probability is significantly lower for SF than for VF ($z = -2.99$, $p = .015$), whereas the difference in response probability between OF and BF did not reach significance ($p > .05$).

Post-hoc interaction contrast analyses showed that the difference in the probability of a correct response between unprocessed and vocoded stimuli is significantly smaller for OF than for SF ($z = -5.20$, $p < .001$) or VF ($z = -7.18$, $p < .001$) and significantly smaller for BF than for SF ($z = -3.50$, $p < .001$) or VF ($z = -5.64$, $p < .001$). The processing contrast is also significantly smaller for SF than for VF ($z = -2.27$, $p = .023$) and significantly smaller for OF than for BF ($z = -2.17$, $p = .03$). The difference in the probability of a correct response between the more proficient adults and the less proficient adolescents is significantly greater for SF compared to VF ($z = 2.84$, $p = .005$), OF ($z = 2.16$, $p = .031$), and BF ($z = 2.61$, $p = .009$), but not significantly different for OF compared to BF, or for VF compared to either OF or BF ($p > .05$).

4. Discussion

The present study used a focus interpretation task with unprocessed and vocoded stimuli to investigate whether the spectro-temporal degradation of the simulated electric speech signal influences how accurately NH native Dutch learners of English of different ages and with different proficiency levels in English interpret prosodically marked focus in the non-native language. Results showed that, as expected, the CI simulation had a significant impact on the focus interpretation accuracy. Listeners were less accurate in identifying the correct context question in the vocoded condition compared to the unprocessed condition. This is in line with previous research on the influence of a CI or CI simulation on the recognition of prosodically focused words in a native language [17, 18, 19, 20, 21, 22]. It reiterates the importance of quality access to prosodic cues for prosody identification [1, 2, 3, 4], regardless of whether listeners are native or non-native listeners.

Overall, the more proficient adults outperformed the less proficient adolescents. Post-hoc analyses showed, however, that the difference between the participant groups was only significant for SF and OF. Even so, this finding is consistent with previous research on the recognition of prosodic focus with non-native listeners with different proficiency levels [15]. It shows that the accuracy of focus recognition increases with increasing proficiency in the non-native language. Note that age could also have played a factor as the more proficient adults were not just more proficient in English and more experienced in English learning in a school setting, but they were also older than the less proficient adolescents. Some studies indicate that the adolescents are of an age where the development in understanding prosodic patterns is expected to be adult-like [34], but others indicate potentially longer developmental periods for prosody-related tasks [35]. However, the fact that the results of the present study are in line with previous research where age did not play a role suggests that the higher accuracy for the more proficient adults can be attributed to the fact that these native Dutch learners of English are the more experienced and more proficient non-native language learners [15, 16]. A more in-depth discussion regarding the covarying age effect will be presented in a forthcoming paper.

The significantly lower probability of a correct response for OF and BF compared to SF and VF observed in both processing conditions and for both participant groups suggests that there might be an underlying cause for this reduced accuracy. Recall that by default the nuclear accent is on the phrase-final word, but that for utterances with prosodically marked linguistic focus, the nuclear accent is on the focused word [8, 6, 9]. The nuclear accent is thus on the object for both the OF and the BF stimuli of this study. That is, the object is the focused word for OF stimuli, but it is also in phrase-final position, which is why the nuclear accent is also on the object for BF. It is therefore possible that the listeners confused OF and BF with one another, resulting in the reduced accuracy for these two focus types. Future research will be dedicated to investigating the interpretation confusions.

Finally, the influence of the CI simulation on the focus interpretation patterns did not significantly differ between the non-native listener groups. Both the less proficient adolescents and the more proficient adults were less accurate in linking focus position to the correct context question in the vocoded condition compared to the unprocessed condition and the more proficient adults tended to outperform the less proficient adolescents regardless of the processing condition. The lack of an interaction here suggests that the difference in the impact of adverse listening conditions between native and non-native listeners [23] cannot be extended to non-native listeners with different proficiency levels. The less proficient adolescents were not more strongly affected by the spectro-temporal degradations than the more proficient adults. Note that the age difference between the non-native listeners cannot be ruled out as an effect, even though some studies indicate that the perception of degraded speech is expected to be adult-like by the age 10–12 years [36]. The covarying age effect will be discussed more thoroughly in a forthcoming paper. Moreover, future research could investigate how the impact of the CI simulation on prosodic focus interpretation for the non-native listeners compares to this impact for native listeners.

5. Conclusion

This study assessed how a CI simulation influences the interpretation of prosodically marked focus in a non-native language. The results confirmed that a CI simulation lowers the accuracy of prosodic focus interpretation and that the accuracy of how well non-native listeners can link prosodic focus position to the correct context question increases with increasing proficiency in the non-native language. It was also revealed that the influence of the spectro-temporally degraded electric speech signal does not interact with the influence of the proficiency level in the non-native language, indicating that less proficient non-native listeners are not more strongly affected by a CI simulation than more proficient non-native listeners.

6. References

[1] D. Başkent, E. Gaudrain, T. N. Tamati, and A. Wagner, "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific foundations of audiology: Perspectives from physics, biology, modeling, and medicine*, A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk, Eds. Plural Publishing Inc., 2016, pp. 285–319.

[2] M. Chatterjee and S.-C. Peng, "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," *Hearing Research*, vol. 235, no. 1-2, pp. 143–156, 2008.

[3] E. Gaudrain and D. Başkent, "Discrimination of voice pitch and vocal-tract length in cochlear implant users," *Ear and Hearing*, vol. 39, no. 2, pp. 226–237, 2018.

[4] M. K. Everhardt, A. Sarampalis, M. Coler, D. Başkent, and W. Lowie, "Meta-analysis on the identification of linguistic and emotional prosody in cochlear implant users and vocoder simulations," *Ear and Hearing*, vol. 41, no. 5, pp. 1092–1102, 2020.

[5] S. Birch and C. Clifton, Jr, "Focus, accent, and argument structure: Effects on language comprehension," *Language and Speech*, vol. 38, no. 4, pp. 365–391, 1995.

[6] J. Cole, "Prosody in context: A review," *Language, Cognition and Neuroscience*, vol. 30, no. 1-2, pp. 1–31, 2015.

[7] P. Welby, "Effects of pitch accent position, type, and status on focus projection," *Language and Speech*, vol. 46, no. 1, pp. 53–81, 2003.

[8] S. Calhoun, "The centrality of metrical structure in signaling information structure: A probabilistic perspective," *Language*, vol. 86, no. 1, pp. 1–42, 2010.

[9] D. R. Ladd, *Intonational phonology*, 2nd ed. Cambridge University Press, 2008.

[10] S. Calhoun, E. Wollum, and E. Kruse Va'ai, "Prosodic prominence and focus: Expectation affects interpretation in Samoan and English," *Language and Speech*, vol. 64, no. 2, pp. 346–380, 2021.

[11] E. Vallduví, "Information structure," in *The Cambridge handbook of formal semantics*, M. Aloni and P. Dekker, Eds. Cambridge University Press, 2016, pp. 728–755.

[12] A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the comprehension of spoken language: A literature review," *Language and Speech*, vol. 40, no. 2, pp. 141–201, 1997.

[13] E. Akker and A. Cutler, "Prosodic cues to semantic structure in native and nonnative listening," *Bilingualism: Language and Cognition*, vol. 6, no. 2, pp. 81–96, 2003.

[14] R. Wayland, C. Guerra, S. Chen, and Y. Zhu, "English focus perception by Mandarin listeners," *Languages*, vol. 4, no. 4, 2019.

[15] R. E. Baker, "Non-native perception of native English prominence," in *Proceedings of the 5th International Conference on Speech Prosody*, 2010, vol. 100171, pp. 1–4.

[16] J. E. Flege, O.-S. Bohn, and S. Jang, "Effects of experience on non-native speakers' production and perception of English vowels," *Journal of Phonetics*, vol. 25, no. 4, pp. 437–470, 1997.

[17] H. Meister, M. Landwehr, V. Pyschny, P. Wagner, and M. Walger, "The perception of sentence stress in cochlear implant recipients," *Ear and Hearing*, vol. 32, no. 4, pp. 459–467, 2011.

[18] H. Meister, M. Landwehr, V. Pyschny, M. Walger, and H. von Wedel, "The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients," *International Journal of Audiology*, vol. 48, no. 1, pp. 38–48, 2009.

[19] R. O'Halpin, "The perception and production of stress and intonation by children with cochlear implants," Ph.D. dissertation, University College London, 2010. [Online]. Available: https://discovery.ucl.ac.uk/id/eprint/20406

[20] C. M. Holt, K. Demuth, and I. Yuen, "The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants," *Ear and Hearing*, vol. 37, no. 4, pp. e256–e262, 2016.

[21] R. T. Kalathottukaren, S. C. Purdy, and E. Ballard, "Prosody perception and musical pitch discrimination in adults using cochlear implants," *International Journal of Audiology*, vol. 54, no. 7, pp. 444–452, 2015.

[22] D. J. van de Velde, N. O. Schiller, V. J. van Heuven, C. C. Levelt, J. van Ginkel, M. Beers, J. J. Briaire, and J. H. M. Frijns, "The perception of emotion and focus prosody with varying acoustic cues in cochlear implant simulations with varying filter slopes," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3349–3363, 2017.

[23] M. L. G. Lecumberri, M. Cooke, and A. Cutler, "Non-native speech perception in adverse conditions: A review," *Speech Communication*, vol. 52, no. 11-12, pp. 864–886, 2010.

[24] Council of Europe, *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing, 2020.

[25] Inspectie van het Onderwijs, "Peil.Engels – Technische rapportage," 2019. [Online]. Available: https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2019/11/08/technisch-rapport-peil.engels/Peil.Engels+TECHNISCHE+RAPPORTAGE.pdf

[26] C. Smits, S. T. Goverts, and J. M. Festen, "The digits-in-noise test: Assessing auditory speech recognition abilities in noise," *The Journal of the Acoustical Society of America*, vol. 133, no. 3, pp. 1693–1706, 2013.

[27] E. Gaudrain, "Vocoder," 2016. [Online]. Available: https://github.com/egaudrain/vocoder

[28] M. K. Everhardt, A. Sarampalis, M. Coler, D. Başkent, and W. Lowie, "Perception of L2 lexical stress in words degraded by a cochlear implant simulation," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Australasian Speech Science and Technology Association Inc., 2019, pp. 102–106.

[29] M. Chatterjee, D. J. Zion, M. L. Deroche, B. A. Burianek, C. J. Limb, A. P. Goren, A. M. Kulkarni, and J. A. Christensen, "Voice emotion recognition by cochlear-implanted children and their normally-hearing peers," *Hearing Research*, vol. 322, pp. 151–162, 2015.

[30] J. R. de Leeuw, "jsPsych: A JavaScript library for creating behavioral experiments in a web browser," *Behavior Research Methods*, vol. 47, no. 1, pp. 1–12, 2015.

[31] K. Lange, S. Kühn, and E. Filevich, ""Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies," *PLoS ONE*, vol. 10, no. 6, 2015.

[32] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[33] R. V. Lenth, "emmeans: Estimated marginal means, aka least-squares means," 2021. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[34] B. Wells, S. Peppé, and N. Goulandris, "Intonation development from five to thirteen," *Journal of Child Language*, vol. 31, no. 4, pp. 749–778, 2004.

[35] L. Nagels, E. Gaudrain, D. Vickers, M. Matos Lopes, P. Hendriks, and D. Başkent, "Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations," *PeerJ*, vol. 8, e8773, 2020.

[36] L. S. Eisenberg, R. V. Shannon, A. Schaefer Martinez, J. Wygonski, and A. Boothroyd, "Speech recognition with reduced spectral cues as a function of age," *The Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2704–2710, 2000.