

University of Groningen

Prosodic Focus Interpretation in Spectrotemporally Degraded Speech by Non-Native Listeners

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

Published in:
Journal of Speech, Language and Hearing Research

DOI:
[10.1044/2023_JSLHR-22-00568](https://doi.org/10.1044/2023_JSLHR-22-00568)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

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

Publication date:
2023

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

Citation for published version (APA):

Everhardt, M. K., Sarampalis, A., Coler, M., Başkent, D., & Lowie, W. (2023). Prosodic Focus Interpretation in Spectrotemporally Degraded Speech by Non-Native Listeners. *Journal of Speech, Language and Hearing Research*, 66(9), 3649-3664. https://doi.org/10.1044/2023_JSLHR-22-00568

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

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.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Research Article

Prosodic Focus Interpretation in Spectrotemporally Degraded Speech by Non-Native Listeners

Marita K. Everhardt,^{a,b,c}  Anastasios Sarampalis,^{c,d}  Matt Coler,^{c,e}  Deniz Başkent,^{b,c,f} 
and Wander Lowie^{a,c} 

^aCenter for Language and Cognition Groningen, University of Groningen, the Netherlands ^bDepartment of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, the Netherlands ^cResearch School of Behavioural and Cognitive Neurosciences, University of Groningen, the Netherlands ^dDepartment of Psychology, University of Groningen, the Netherlands ^eCampus Fryslân, University of Groningen, the Netherlands ^fW.J. Kolff Institute for Biomedical Engineering and Materials Science, University Medical Center Groningen, University of Groningen, the Netherlands

ARTICLE INFO**Article History:**

Received September 29, 2022

Revision received March 26, 2023

Accepted May 19, 2023

Editor-in-Chief: Peggy B. Nelson

Editor: Yi Shen

https://doi.org/10.1044/2023_JSLHR-22-00568

ABSTRACT

Purpose: This study assesses how spectrotemporal degradations that can occur in the sound transmission of a cochlear implant (CI) may influence the ability of non-native listeners to recognize the intended meaning of utterances based on the position of the prosodically focused word. Previous research suggests that perceptual accuracy and listening effort are negatively affected by CI processing (or CI simulations) or when the speech is presented in a non-native language, in a number of tasks and circumstances. How these two factors interact to affect prosodic focus interpretation, however, remains unclear.

Method: In an online experiment, normal-hearing (NH) adolescent and adult native Dutch learners of English and a small control group of NH native English adolescents listened to CI-simulated (eight-channel noise-band vocoded) and non-CI-simulated English sentences differing in prosodically marked focus. For assessing perceptual accuracy, listeners had to indicate which of four possible context questions the speaker answered. For assessing listening effort, a dual-task paradigm was used with a secondary free recall task.

Results: The results indicated that prosodic focus interpretation was significantly less accurate in the CI-simulated condition compared with the non-CI-simulated condition but that listening effort was not increased. Moreover, there was no interaction between the influence of the degraded CI-simulated speech signal and listening groups in either their perceptual accuracy or listening effort.

Conclusion: Non-native listeners are not more strongly affected by spectrotemporal degradations than native listeners, and less proficient non-native listeners are not more strongly affected by these degradations than more proficient non-native listeners.

Cochlear implants (CIs) are auditory prostheses that can partially restore hearing in individuals with profound sensorineural hearing loss. An electrode array inserted in the cochlea sends electrical signals directly to the auditory nerve. This electric stimulation makes speech perception possible, yet a speech signal transmitted this way in electric hearing has much less fine spectrotemporal detail than

an acoustic speech signal transmitted in normal hearing (for details on electric hearing, see Başkent et al., 2016). The spectrotemporal degradation obscures those qualities of the speech signal that function as cues to linguistic phenomena (Moberly et al., 2021; Pisoni, 2005; Shannon et al., 2004). This is particularly the case for fundamental frequency (f_0), a cue mainly related to the pitch of a speaker's voice and, in a broader sense, to prosodic features of speech (Chatterjee & Peng, 2008; Gaudrain & Başkent, 2018). Limited access to the prosodic cue f_0 in electric hearing has been shown to compromise the ability of listeners to recognize prosodic patterns in the native

Correspondence to Marita K. Everhardt: m.k.everhardt@rug.nl. **Disclosure:** The authors have declared that no competing financial or non-financial interests existed at the time of publication.

language (for an overview, see Everhardt et al., 2020). This study explores how spectrotemporal degradations similar to those that can occur in electric hearing through a CI simulation (i.e., an acoustic approximation of electric hearing implemented by a noise-band vocoder) influences the recognition of the intended meaning 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, important, or in contrast with what was previously said (Birch & Clifton, 1995; Bishop, 2012; Cole, 2015; Wayland et al., 2019; Welby, 2003). In English, linguistic focus is prosodically marked by a *pitch accent*, specifically by the *nuclear accent* (i.e., the head of the intonational phrase). Pitch accents are realized with localized phonetic prominence relative to the words around it, generally indicated not only by higher f_0 and greater f_0 movement but also by increased intensity and increased duration. By default, the word at the end of an intonational phrase carries the nuclear accent in English, unless a preceding word is in focus, in which case the word in focus carries the nuclear accent (Calhoun, 2010; Cole, 2015; Ladd, 2008). The context of the discourse influences what part of an utterance is in focus (Calhoun et al., 2021; Vallduví, 2016). Consider the following example (note: [...] _F indicates the focus and capitalization indicates the nuclear accent):

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

The focus patterns in 1a–1d are in response to the context questions that precede each one. 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 recognize focused words and the discourse implications of the prosodically marked linguistic focus, which facilitates sentence comprehension (for a review, see Cutler et al., 1997). For instance, a recent study by Calhoun et al. (2021) has shown that native listeners can correctly recognize the intended meaning of an utterance based on the position of the

prosodically focused word. In the experiment, adult native (New Zealand) English listeners were presented with canonical English sentences with the nuclear accent either on the subject or on the object. Listeners were asked to indicate which of two context questions was the most likely question given the way the speaker responds. These context questions were designed to prompt either the SF response or 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 pattern and were able to identify the correct context question.

Non-native listeners, however, recognize prosodic patterns less accurately and less efficiently (e.g., Akker & Cutler, 2003; Wayland et al., 2019). Note that we make no specific distinction in our discussion of the literature or in our own study between learners of a *second language* (who use the target language in daily life as they, e.g., live in a country where this language is spoken) and learners of a *foreign language* (who learn the target language in a classroom setting and generally do not use this language in daily life). Instead, the umbrella term *non-native language learner* or *non-native listener* is used throughout. A study by Baker (2010) specifically showed that non-native listeners performed significantly less accurately compared with native listeners during a task where they had to indicate whether the prosodically marked linguistic focus pattern of a response was appropriate given a context question for matched and mismatched question–answer pairs in English. The adult native (American) English listeners were very likely to accept matched pairs and reject mismatched pairs, thus accurately linking the focus pattern to the correct context questions. The adult Korean and Mandarin listeners, in contrast, had greater difficulty interpreting the prosodically marked linguistic focus, mainly evidenced by the lower accuracy in rejecting mismatched question–answer pairs. The study revealed a crucial finding that as the Korean and Mandarin listeners' English proficiency increased, as assessed through the Versant English Test (<http://www.ordinate.com/products/english.jsp>), their accuracy also improved. This is in line with previous research on the perception of non-native speech in the segmental domain showing a beneficial effect of the amount of experience with the non-native language (e.g., Flege et al., 1997). These studies show that non-native language processing—including prosodic focus interpretation—improves with increasing experience and proficiency in the non-native language.

Little is known about how prosodic focus interpretation for non-native language learners may be affected by the spectrotemporal degradations inherent to electric hearing. 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 with their normal-hearing (NH) peers (Meister et al., 2009, 2011). Similar findings were reported across several studies with different CI populations. Early deafened and implanted children (O’Halpin, 2010), prelingually deafened adolescent CI users (Holt et al., 2016), and adults who were deafened either prelingually or postlingually but implanted late (Kalathottukaren et al., 2015) all showed deficits compared with their NH peers. Additionally, NH adults showed similar deficits in prosodic focus interpretation, when listening to CI-simulated (vs. unprocessed) speech (van de Velde et al., 2017). It remains unclear how such spectrotemporal degradations could influence the interpretation of prosodically marked linguistic focus in a non-native language. In terms of generic non-native language skills in CI users, two cohort studies by Beeres-Scheenstra et al. (2017, 2020) showed that Swiss German 10- to 18-year-old adolescent CI users have difficulty performing listening tasks in the non-native language. That is, only a low percentage of the CI users reached the Swiss school norm for listening skills in either English or French as a non-native language, showing a clear disadvantage for the non-native CI users compared with their non-native NH peers. For NH listeners listening to CI-simulated speech, a recent study by Yang et al. (2022) showed that adult native Mandarin learners of English living in the United States have greater difficulty recognizing English consonants and vowels as well as sentences in noise-band vocoded stimuli than their native English peers. Increasing the number of channels improved phoneme and sentence recognition for both listening groups, yet the native listeners showed a greater amount of improvement as the number of channels increased. The lower recognition accuracy of the non-native listeners across all conditions except for the two-channel condition can predominantly be attributed to a non-native disadvantage, which becomes more prominent as the number of channels increases. These results thus indicate that the non-native disadvantage is less prominent for the more adverse conditions, whereas speech-in-noise perception studies indicate that non-native listeners are more strongly affected by adverse listening conditions than native listeners (for a review, see Lecumberri et al., 2010).

Going beyond perceptual accuracy, adverse listening conditions have also been shown to greatly impact listening effort, especially for non-native listeners (Borghini & Hazan, 2018; Peng & Wang, 2019). Listening effort refers to the mental effort that occurs when performing a task that involves listening, where listeners deliberately allocate mental resources to the execution of a task (Pichora-Fuller et al., 2016). It is based on the notion that listeners have a limited cognitive capacity that can be allocated when

performing tasks (Broadbent, 1958; Kahneman, 1973). The more cognitively demanding the task, the more mental resources are required to execute the task. When performing multiple tasks simultaneously, there is a competition for the mental resources that can be allocated. Listening effort can in this case be assessed using a dual-task paradigm where listeners perform a primary and a secondary task simultaneously. A cognitively demanding primary task requires a high proportion of mental effort, limiting the allocation of mental resources to the execution of the simultaneously performed task. That is, fewer mental resources are available for the secondary task, resulting in a decrease in performance. Lower performance in the secondary task is thus accepted as being indicative of increased listening effort for the primary task.

Studies using the dual-task paradigm have shown that listening effort increases in adverse listening conditions, sometimes even when accuracy may not reflect a measurable change. For example, Sarampalis et al. (2009) showed that during a speech-in-noise perception task, the performance in the secondary task improved when noise reduction algorithms were applied, whereas the performance in the primary task did not differ between the noise reduction processing conditions. In other words, at low signal-to-noise ratios, reducing noise by applying algorithms similar to those used in hearing aids decreases listening effort but does not modulate perceptual accuracy. For electric hearing, Pals et al. (2013) showed that increasing the spectral resolution improves perceptual accuracy and reduces listening effort. In this study, listeners were presented with noise-band vocoded stimuli that varied from two up to 24 channels. The results showed that the performance in both the primary task and the secondary task significantly improved by increasing the number of channels (up to six channels for the primary task and up to eight channels for the secondary task). Listening effort and perceptual accuracy are thus both modulated by the spectral resolution of a CI-simulated speech signal.

Related to non-native listening, listening effort is influenced not only by adverse listening conditions but also by listener characteristics (Pichora-Fuller et al., 2016). For instance, Peng and Wang (2019) looked at differences in listening effort between a group of adult native (American) English listeners and two groups of adult non-native English listeners with either Mandarin Chinese as their native language or another native language (e.g., Hindu, Korean, and Portuguese). In the study, listening effort in a range of acoustic conditions (mimicking classroom environments) was measured objectively using a dual-task paradigm, with a primary speech comprehension task in English and a secondary adaptive pursuit rotor dot-tracing task, as well as subjectively using questionnaires. All listeners experienced comprehension deficits in more

adverse listening conditions. Across acoustic conditions, the group of non-native listeners with a native language other than Chinese scored significantly lower in the secondary task compared with the other two groups, indicating increased listening effort for this group of non-native listeners. Moreover, the questionnaire data revealed an increase in self-reported listening effort under more adverse acoustic conditions for all listening groups, with both groups of non-native listeners reporting even more listening effort than the native listeners. This subjective measure thus suggests that non-native listeners are more strongly affected by adverse conditions than native listeners in terms of listening effort. In another study, Borghini and Hazan (2018) assessed listening effort for non-native listeners using pupillometry. The pupil responses of adult native (British) English listeners and adult native Italian learners of English were measured during a speech intelligibility task in quiet as well as in babble background noise for which the intelligibility performance level was matched across groups. The results showed that the pupil dilation was significantly greater for the non-native listeners compared with the native listeners, both in quiet and in the babble background noise conditions. In a follow-up study, Borghini and Hazan (2020) corroborated that non-native listeners require a greater listening effort than native listeners to perform a speech intelligibility task in adverse listening conditions, even when overall intelligibility is matched. This study also showed that native Italian learners of English with a higher proficiency level in English, determined by their International English Language Testing System Listening score, could tolerate background noise to a greater degree than Italian listeners with a lower English proficiency level. That is, more proficient non-native listeners could better ignore interfering noise. The non-native proficiency level, however, did not have an influence on pupil dilation. This suggests that listening effort was not reduced with increasing proficiency in the non-native language. However, the authors also note that pupillometry might not be the most reliable technique to uncover individual differences. As such, whether listening effort may be modulated by the proficiency level in the non-native language remains unclear.

This study assesses in an online experiment how spectrotemporal degradations (via CI simulations) could influence the way non-native listeners with different proficiency levels interpret prosodically marked linguistic focus. The study also assesses how these degradations could influence listening effort. Based on the CI literature, we predict that listeners will interpret prosodic focus less accurately in the CI-simulated condition than in the non-CI-simulated condition (e.g., van de Velde et al., 2017) and that listening effort will be increased for the

CI-simulated condition (e.g., Pals et al., 2013). Based on the literature with non-native listeners, we predict that non-native listeners will interpret prosodic focus less accurately than native listeners (e.g., Wayland et al., 2019) and that focus processing is more effortful for the non-native listeners than for the native listeners (e.g., Borghini & Hazan, 2018, 2020). Furthermore, we expect that prosodic focus interpretation accuracy will be higher for the more proficient non-native listeners than for the less proficient non-native listeners (e.g., Baker, 2010). Whether listening effort will be modulated by the proficiency level in the non-native language remains unclear. Central to our study is the question whether the non-native listeners will be more strongly affected by the spectrotemporal degradations than the native listeners and whether the less proficient non-native listeners will, in turn, be more strongly affected than the more proficient non-native listeners. This interaction has—to our knowledge—never been studied before for the interpretation of prosodically marked linguistic focus in CI-simulated speech.

Method

Participants

Three groups took part in this online study: two groups of native Dutch learners of English (also reported in Everhardt et al., 2022) and a small control group of native English listeners. The two groups of native Dutch participants were included as a simple investigation into the influence of the proficiency level in the non-native language, using age and experience as a crude measure of non-native language proficiency. The less proficient non-native listeners included eighteen 12- to 14-year-old native Dutch secondary school students ($M_{\text{age}} = 13.5 \pm 0.93$) recruited through secondary schools in the North of the Netherlands and the personal networks of the authors. These adolescents had a limited amount of experience with English learning in a school setting and an estimated average Common European Framework of Reference for Languages (CEFR; Council of Europe, 2020) level of English listening skills of A2 (Inspectie van het Onderwijs, 2019). The more proficient non-native listeners included thirty-two 18- to 30-year-old native Dutch first-year psychology students ($M_{\text{age}} = 20.6 \pm 2.24$) from the University of Groningen recruited through the psychology participant pool of the University of Groningen. These adults had more years of experience with English learning in a school setting and 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. The small control group included four 13- to 14-year-old native English secondary school

students ($M_{\text{age}} = 13.5 \pm 0.4$), recruited through the personal networks of the authors.

The adolescents in this study are of an age where the development in understanding prosodic patterns is expected to be adultlike (Wells et al., 2004). Exclusion criteria for all participants were self-reported learning disabilities or language impairments. Moreover, the participants were required to have normal hearing and normal (or corrected-to-normal) vision. To confirm normal hearing, participants were asked to indicate that they have no known hearing problems. In addition, they were requested to successfully complete either the Dutch or the English version of the online digits-in-noise (DIN) test (Smits et al., 2013, 2016) before participation in the online study. This test is a validated instrument that assesses speech recognition in noise and is used in the Netherlands as a shorthand proxy for assessing hearing ability. All participants declared that they passed the online DIN test.

The study and data collection for the native and non-native adolescents was approved by the Research Ethics Review Committee of the Faculty of Arts, University of Groningen (CETO 70647971). In order to recruit the non-native adults, a separate study protocol was necessary. This protocol was approved by the Ethics Committee of Psychology of the University of Groningen (ECP PSY-2021-S-0048). Written informed consent was obtained from all participants and the parents/guardians of the adolescents before participation in the study. The native adolescents volunteered their time, the non-native adolescents were paid for their time, and the non-native adults received course credit for participation.

Stimuli

Twenty-six English sentences were constructed (24 trial items and two practice items) with a subject–verb–article–object structure such as “John kicked the ball.” The subjects of the trial items were 12 female first names (six monosyllabic and six disyllabic) and 12 male first names (six monosyllabic and six disyllabic), the verbs were transitive monosyllabic verbs in the past tense, and the objects were singular common nouns (12 monosyllabic and 12 disyllabic) preceded by the definite article “the.” The subjects, verbs, and objects were selected from the 1,000 most frequent singular proper nouns, the 1,000 most frequent past tense lexical verbs, and the 1,000 most frequent singular common nouns (preceded by “the”) respectively according to the Corpus of Contemporary American English (<https://www.english-corpora.org/coca/>) and the British National Corpus (<https://www.english-corpora.org/bnc/>). The practice items were similarly constructed. Four context questions were created for each trial and practice item. These questions were designed to elicit an 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 to ensure speaker variability, mimicking a naturalistic setting. Each speaker recorded the 24 trial items and two practice items in all focus forms. The speakers were instructed to produce each sentence at least 3 times and as naturally as possible. 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. For each speaker, the most natural production of each SF, VF, OF, and BF sentence was selected based on auditory judgments by a trained linguist (the first author), only selecting sentences with the correct focus pattern that were not overemphasized. If all productions of a sentence were produced with a natural focus pattern, selection was based on the clarity of the recording (e.g., no obvious background noise). The 96 trial items (24 sentences \times 4 focus types) were subsequently 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 sentences to the stimuli set. The practice items were similarly divided.

The selected trial items were acoustically analyzed with respect to differences in peak f_0 , peak intensity, and duration. Using a Praat script (Version 6.2.14; Boersma & Weenink, 2022), the f_0 , intensity, and duration measures were extracted for the stressed syllable of each subject, verb, and object. These measures were subsequently analyzed by fitting linear mixed-effects models (LMMs) in the R environment (Version 4.2.1) using the *lmer* function of the *lme4* package (Version 1.1-29; Bates et al., 2015). For each measure, the LMM included an interaction between *constituent* (subject, verb, and object) and *focus* (SF, VF, OF, and BF) as well as a by-speaker random intercept. Post hoc pairwise comparisons were carried out using the *emmeans* package (Version 1.7.0; Lenth, 2022). The pairwise comparisons in Table 1 show that constituents in focus have higher peak f_0 and higher peak intensity than constituents that are not in focus within that sentence (note that a direct comparison of constituent duration within a sentence is not a reliable measure of phonetic prominence due to the fact that constituents differed in word length). For example, in SF stimuli, the f_0 and intensity measures are significantly higher for the subject (vs. verb or object). In BF stimuli, the f_0 and intensity measures are lower for the object (vs. subject or verb), despite the fact that the object carries the nuclear accent. In fact, the subject has a significantly higher peak intensity than the object and also a higher peak f_0 , but this difference

Table 1. Differences in peak fundamental frequency (f_0), peak intensity, and duration between relevant constituents (subject, verb, and object) for each focus type (subject focus [SF], verb focus [VF], object focus [OF], and broad focus [BF]).

Constituent comparison per focus type	f_0 (semitones)			Intensity (dB)			Duration (ms)		
	Est.	SE	<i>t</i>	Est.	SE	<i>t</i>	Est.	SE	<i>t</i>
SF									
Subject vs. verb	4.84	0.84	5.76***	7.71	0.72	10.78***	-11.1	25.8	-0.43
Subject vs. object	6.99	0.83	8.42***	10.26	0.72	14.35***	-17.5	25.8	-0.68
VF									
Verb vs. subject	2.90	0.82	3.54***	3.03	0.72	4.25***	121.5	25.8	4.71***
Verb vs. object	5.40	0.83	6.51***	9.12	0.72	12.76***	41.6	25.8	1.61
OF									
Object vs. subject	1.30	0.82	1.58**	1.45	0.72	2.03	105.1	25.8	4.08***
Object vs. verb	2.44	0.85	2.87*	2.81	0.72	3.93***	63.4	25.8	2.46*
BF									
Object vs. subject	-1.40	0.81	-1.72	-4.42	0.72	-6.18***	69.4	25.8	2.69*
Object vs. verb	-0.48	0.82	-0.58	-1.49	0.72	-2.08	36.1	25.8	1.40

Note. Est. = estimate; SE = standard error.

* $p < .05$. ** $p < .01$. *** $p < .001$.

was not significant. This shows that speakers produced the subject with higher absolute phonetic prominence than the nuclear accented object, indicating that the nuclear accent is downstepped in BF stimuli. The pairwise comparisons in Table 2 show that constituents have a higher peak f_0 , higher peak intensity, and longer duration when they are in focus. For example, the f_0 , intensity, and duration measures for the verb are significantly higher in VF stimuli (vs. SF, OF, or BF stimuli). Note that the difference in duration is only a significant indicator of phonetic prominence for the verbs. Moreover, SF and BF stimuli did not differ in the peak intensity of the subject, indicating that the intensity of the subject was similarly prominent for

SF and BF stimuli. An example of the f_0 contours, the intensity, and the duration for the SF, VF, OF, and BF stimuli can be found in Figure 1 (left column).

The selected stimuli discussed and analyzed above were manipulated for inclusion in the experiment. Acoustic CI simulations of these stimuli were created by means of a vocoder (Version 1.0; Gaudrain, 2016) implemented in MATLAB (R2018a). Vcoded stimuli were created using an eight-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

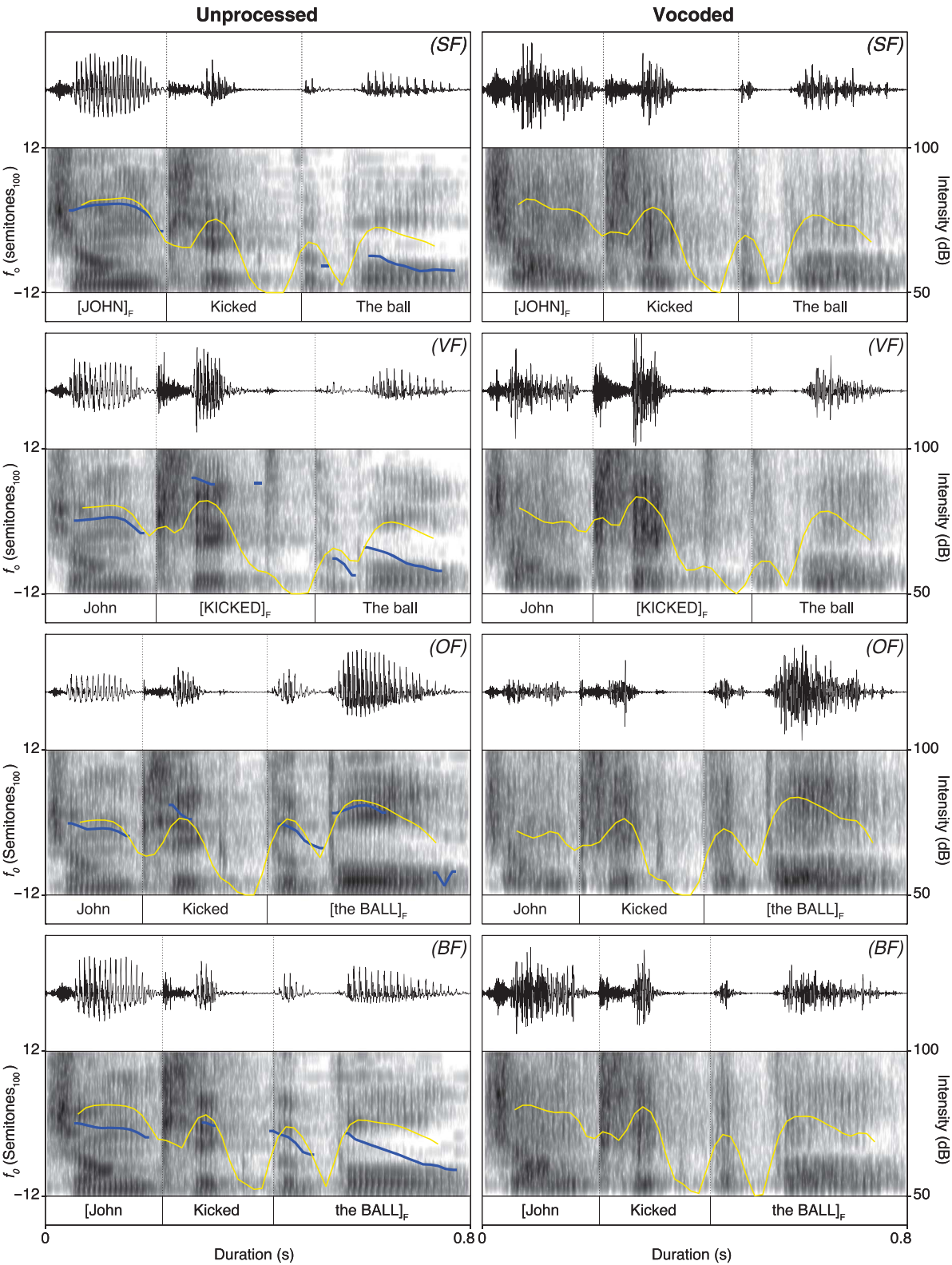
Table 2. Differences in peak fundamental frequency (f_0), peak intensity, and duration between relevant focus types (subject focus [SF], verb focus [VF], object focus [OF], and broad focus [BF]) for each constituent (subject, verb, and object).

Focus type comparison per constituent	f_0 (semitones)			Intensity (dB)			Duration (ms)		
	Est.	SE	<i>t</i>	Est.	SE	<i>t</i>	Est.	SE	<i>t</i>
Subject									
SF vs. VF	2.92	0.81	3.59**	3.98	0.72	5.58***	42.5	25.8	1.65
SF vs. OF	2.69	0.81	3.31**	3.80	0.72	5.32***	38.0	25.8	1.47
SF vs. BF	2.30	0.81	2.83*	1.76	0.72	2.46***	32.6	25.8	1.27
Verb									
VF vs. SF	4.82	0.85	5.68***	6.76	0.72	9.45***	67.9	25.8	2.64*
VF vs. OF	3.82	0.85	4.50***	4.21	0.72	5.89***	75.3	25.8	2.92*
VF vs. BF	3.21	0.83	3.87**	3.74	0.72	5.23***	78.3	25.8	3.04*
Object									
OF vs. SF	5.59	0.84	6.67***	7.90	0.72	11.06***	49.6	25.8	1.92
OF vs. VF	4.02	0.83	4.85***	7.72	0.72	10.80***	29.7	25.8	1.15
OF vs. BF	2.30	0.82	2.81*	3.82	0.72	5.35***	30.3	25.8	1.18

Note. Est. = estimate; SE = standard error.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 1. Example stimuli showing the fundamental frequency (f_0) contours, the intensity, and the duration for the subject focus (SF), verb focus (VF), object focus (OF), and broad focus (BF) stimuli (left column) and for the subsequently vocoded SF, VF, OF, and BF stimuli (right column). Blue lines show the f_0 contours. Yellow lines show the intensity.



extracted by half-wave rectification and low-pass filtering at a cutoff of 160 Hz using a zero-phase fourth-order Butterworth filter. These parameters resembled those of a previous study (Everhardt et al., 2019) but were modified to more closely approximate the perceptual abilities of the average CI listener by selecting an eight-channel vocoder and a cutoff frequency of 160 Hz (e.g., Chatterjee & Peng, 2008, 2015). 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. The influence of the CI simulations on the stimuli is visualized in Figure 1, showing the absence of f_0 contours (undetected by Praat) and a reduction in the spectrotemporal detail for the vocoded stimuli (right column). Both the unprocessed (i.e., nonvocoded) and vocoded stimuli were included in the final stimuli set, resulting in a total of 192 stimuli (24 sentences \times 4 focus types \times 2 processing strategies) that were used in the focus interpretation task discussed below.

Procedure

Participants completed an online experiment assessing the influence of the CI simulation on the processing of prosodically marked linguistic focus in English sentences in both a single-task and a dual-task condition. Focus interpretation accuracy was assessed using a task that tests whether listeners could link the focus pattern to the correct context question (for a similar design, see Calhoun et al., 2021). Listening effort was assessed using a dual-task paradigm (for a similar design, see Sarampalis et al., 2009). The experiment was coded in jsPsych (Version 6.1.0; de Leeuw, 2015), and data collection was managed through a JATOS (Just Another Tool for Online Studies) server (Version 3.5.5; Lange et al., 2015).

The primary task of the experiment (used in both the single-task and the dual-task paradigm) was a single-interval four-alternative forced-choice (1I-4AFC) focus interpretation task with unprocessed 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 randomized order with the constraint that immediate succession of same-sentence stimuli, regardless of the focus pattern, would not be possible.

The secondary task (used in the dual-task paradigm only) was a free recall task, performed simultaneously with the primary task. Participants were instructed to

remember the names that were mentioned in the sentences of the primary task. After every six sentences, participants were asked to write down as many names of the previous six sentences as possible, in any order they preferred. The stimuli in the dual-task condition were presented in randomized order with the constraint that same-sentence stimuli, regardless of the focus pattern, could not occur within a free recall span of six trials. This ensured that each span of six recall trials contained unique sentences and thus unique names that had to be remembered.

The experiment was divided into four blocks, one for each processing condition per task paradigm. The 96 trial items were split into two equal lists. One list with 48 stimuli (12 sentences \times 4 focus types) was used in both the single-task paradigm block with unprocessed stimuli and the dual-task paradigm block with vocoded stimuli. The other list with 48 stimuli (12 sentences \times 4 focus types) was used in both the dual-task paradigm block with unprocessed stimuli and the single-task paradigm block with vocoded stimuli. Each list contained six sentences with female first names (three monosyllabic and three disyllabic) and six sentences with male first names (three monosyllabic and three disyllabic). Moreover, the stimuli were evenly divided between the two lists such that each speaker contributed a total of three SF, three VF, three OF, and three BF sentences to each stimuli list.

The blocks were presented in pseudo-randomized order; participants were presented either with the two blocks with unprocessed stimuli first and the two blocks with vocoded stimuli next or vice versa. The order of the task paradigms was also counterbalanced but was identical for both processing conditions. That is, if participants were presented first with the single-task paradigm and then with the dual-task paradigm in the unprocessed condition, they would also be presented with the single-task paradigm first and the dual-task paradigm next in the vocoded condition. 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 familiarize themselves with vocoded speech.

Results

The response data of the primary and secondary tasks were analyzed by fitting generalized linear mixed-effects models (GLMMs) in the R environment (Version 4.2.1) using the *glmer* function of the *lme4* package (Version 1.1-29; Bates et al., 2015). The prosodic focus interpretation patterns of the primary task were fitted as the estimated probability of a correct response. The free recall performance of the secondary task was fitted as the estimated probability of a correctly recalled name. The need for predictor variables and by-participant and by-stimulus random slopes for necessary predictors was assessed through stepwise model comparisons using the *anova* function, starting from a basic model with only by-participant and by-stimulus random intercepts. The final model for the primary task included an interaction between *processing* (unprocessed and vocoded) and *focus* (SF, VF, OF, and BF), an interaction between *group* (native adolescents, non-native adults, and non-native adolescents) and *focus*, and an interaction between *paradigm* (single-task and dual-task) and *focus*. This model also included by-participant and by-stimulus random slopes for *processing* and *focus*. The final model for the secondary task only included the predictor variable *recall trial* (1–6) as the model comparisons showed that it was not necessary to add *processing*, *group*, or *focus*.

This model also included by-participant random slopes for *recall trial* and a by-stimulus random intercept. Model criticism was applied to the final models of the primary and secondary tasks by excluding observations with residuals > 2.5 SDs from the mean. Post hoc analyses were performed and visualized using the *emmeans* package (Version 1.7.0; Lenth, 2022). The data and code are available at <https://doi.org/10.34894/HCNIE8>.

The model predictions for the primary task are visualized in Figure 2, showing the interaction between *processing* and *focus* (Panel A), the interaction between *group* and *focus* (Panel B), and the interaction between *paradigm* and *focus* (Panel C). The post hoc pairwise comparisons shown in Table 3 revealed that the probability of a correct response is significantly higher for unprocessed (vs. vocoded) stimuli for all focus types. This processing contrast was more prominent for some focus types than for others, but these differences are beyond the scope of this study. Moreover, the response probability is significantly higher for the non-native adults compared with the non-native adolescents for SF and OF, but not for VF or BF. The post hoc analyses revealed no significant differences between the native adolescents and either the non-native adults or the non-native adolescents. The difference in response probability between the single-task and the dual-

Figure 2. Probability of a correct response per focus type (subject focus [SF], verb focus [VF], object focus [OF], and broad focus [BF]) for unprocessed (red triangles) and vocoded (blue circles) stimuli averaged across task paradigms and participant groups (Panel A); for the native adolescents (orange asterisks), non-native adults (purple diamonds), and non-native adolescents (green squares) averaged across processing conditions and task paradigms (Panel B); and for the single-task (pink inverted triangles) and dual-task (black stars) paradigms averaged across processing conditions and participant groups (Panel C). Error bars show 95% confidence intervals.

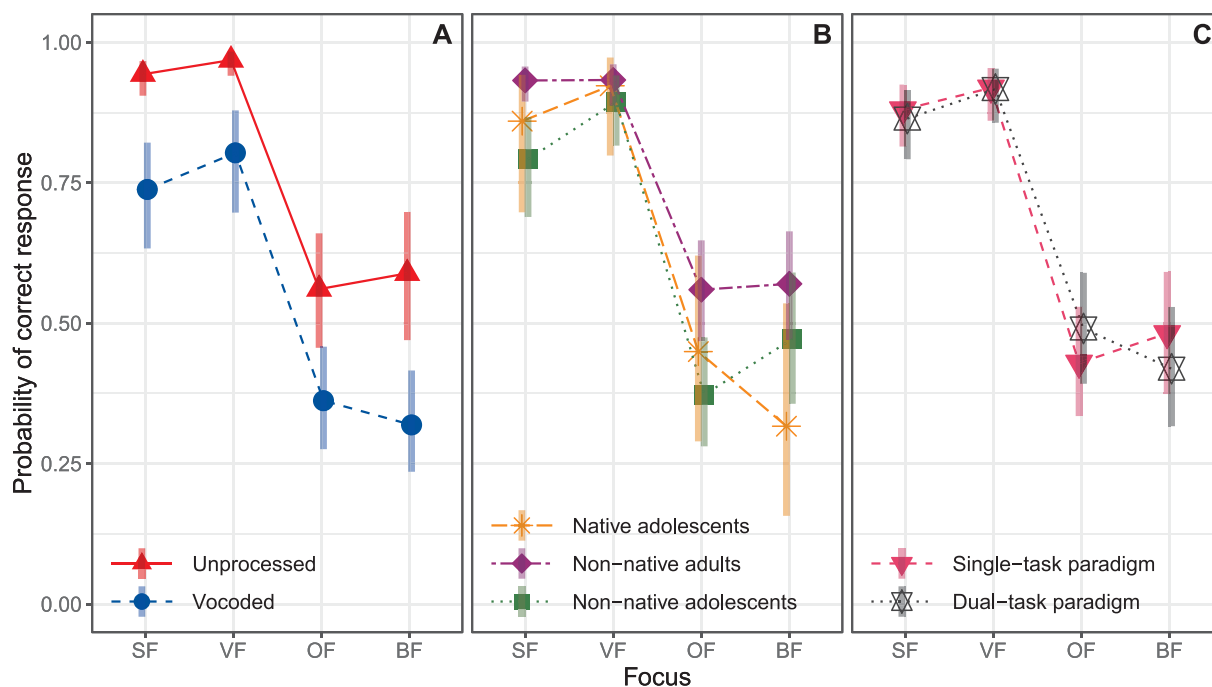


Table 3. Differences in the probability of a correct response between processing conditions (unprocessed and vocoded), task paradigms (single-task and dual-task), and participant groups (native adolescents, non-native adults, and non-native adolescents) for each focus type (subject focus [SF], verb focus [VF], object focus [OF], and broad focus [BF]).

Processing/group/focus type comparison	SF			VF		
	OR	SE	z	OR	SE	z
Unprocessed vs. vocoded	5.89	1.09	9.61***	7.49	1.46	10.31***
Native adolescents vs. non-native adults	0.45	0.22	-1.61	0.86	0.47	-0.28
Native adolescents vs. non-native adolescents	1.61	0.82	0.93	1.42	0.80	0.62
Non-native adults vs. non-native adolescents	3.61	1.01	4.59***	1.65	0.50	1.66
Single-task vs. dual-task	1.15	0.18	0.92	1.03	0.17	0.17
	OF			BF		
	OR	SE	z	OR	SE	z
Unprocessed vs. vocoded	2.25	0.34	5.35***	3.06	0.47	7.29***
Native adolescents vs. non-native adults	0.64	0.22	-1.29	0.35	0.17	-2.22
Native adolescents vs. non-native adolescents	1.37	0.49	0.90	0.52	0.25	-1.35
Non-native adults vs. non-native adolescents	2.14	0.41	3.94***	1.48	0.38	1.53
Single-task vs. dual-task	0.78	0.10	-1.95	1.29	0.17	1.98*

Note. OR = odds ratio; SE = standard error.

* $p < .05$. *** $p < .001$.

task paradigm was marginally significant for BF, showing a higher response probability for the single-task (vs. dual-task) paradigm, but did not reach significance for any of the other focus types (note that the paradigm difference for the primary task will not be discussed any further, since we are mainly interested in the results of the secondary task for the dual-task paradigm). The model predictions in Figure 2 also show that the probability of a correct response is relatively high for SF and VF, but much lower for OF and BF. The post hoc pairwise comparisons shown in Table 4 confirmed that the response probability is significantly lower for OF (vs. SF or VF) and BF (vs. SF or VF) for all processing conditions, participant groups, and task paradigms. No significant differences in response probability between BF and OF were found, and the probability of a correct response was only significantly lower for SF (vs. VF) for the non-native adolescents.

The significantly lower probability of a correct response for OF and BF stimuli compared with SF and VF stimuli for all processing conditions, participant groups, and task paradigms suggests that there might be an underlying cause for this reduced accuracy. Looking at the incorrect responses in addition to the correct responses could provide insight into the prosodic focus interpretation strategies for these specific focus types. Figure 3 shows the focus interpretation matrix as the proportion of SF, VF, OF, and BF responses per focus type for each processing condition, participant group, and task paradigm. It can be observed that SF and VF stimuli are hardly ever confused with other focus types. Since the proportion of correct responses is much lower for OF and BF stimuli across conditions, these focus types are more

frequently confused with other focus types. The matrix shows that OF stimuli are most frequently confused with BF stimuli and that BF stimuli are most frequently confused with SF stimuli.

The GLMM predictions for the secondary task are provided in Table 5, showing the main effect for *recall trial* (bear in mind that the final model for the secondary task did not include any other predictor variables as model comparisons showed that this was not necessary). These predictions indicate that the probability of a correctly recalled name is higher for names at the end of a recall block compared with names in the middle of a block. This is known as the recency effect, indicating that the last few items are stored in short-term memory (Glanzer & Cunitz, 1966). The predictions also indicate that the accuracy is higher for names at the start of a block compared with names in the middle of a block. This is known as the primacy effect, indicating that the first few items are stored in long-term memory (Atkinson & Shiffrin, 1968). The post hoc pairwise comparisons shown in Table 6 confirmed the recency effect, as the probability of a correctly recalled name is significantly higher for the last two recall trials (Recall Trials 5 and 6) compared with earlier recall trials. The post hoc analysis also confirmed the primacy effect, as the recall probability is significantly higher for the first recall trial compared with the third recall trial.

Discussion

This study used a prosodic focus interpretation task with CI-simulated stimuli to investigate how the

Table 4. Differences in the probability of a correct response between focus types (subject focus [SF], verb focus [VF], object focus [OF], and broad focus [BF]) for each processing condition (unprocessed and vocoded), participant group (native adolescents, non-native adults, and non-native adolescents), and task paradigm (single-task and dual-task).

Focus type comparison				Unprocessed			Vocoded		
				OR	SE	z	OR	SE	z
BF vs. SF				0.09	0.03	−7.46***	0.17	0.05	−5.75***
BF vs. VF				0.05	0.02	−8.77***	0.11	0.04	−6.60***
BF vs. OF				1.12	0.32	0.40	0.83	0.24	−0.67
SF vs. VF				0.54	0.17	−1.95	0.69	0.19	−1.34
SF vs. OF				13.00	3.89	8.57***	4.96	1.40	5.69***
VF vs. OF				23.97	8.73	8.72***	7.19	2.47	5.74***
	Native adolescents			Non-native adults			Non-native adolescents		
	OR	SE	z	OR	SE	z	OR	SE	z
BF vs. SF	0.08	0.05	−4.27***	0.10	0.03	−7.78***	0.23	0.08	−4.30***
BF vs. VF	0.04	0.02	−5.75***	0.10	0.03	−7.42***	0.11	0.04	−6.50***
BF vs. OF	0.57	0.31	−1.02	1.04	0.28	0.16	1.50	0.47	1.32
SF vs. VF	0.51	0.25	−1.36	0.99	0.27	−0.95	0.45	0.13	−2.71*
SF vs. OF	7.50	3.48	4.34***	10.79	2.99	8.59***	6.40	1.88	6.32***
VF vs. OF	14.61	8.48	4.62***	10.94	3.66	7.15***	14.17	5.18	7.26***
				Single-task			Dual-task		
				OR	SE	z	OR	SE	z
BF vs. SF				0.13	0.04	−6.49***	0.11	0.04	−6.83***
BF vs. VF				0.08	0.03	−7.42***	0.07	0.02	−8.08***
BF vs. OF				1.24	0.36	0.74	0.75	0.22	−1.01
SF vs. VF				0.65	0.19	−1.48	0.58	0.17	−1.87
SF vs. OF				9.77	2.82	7.89***	6.61	1.91	6.54***
VF vs. OF				15.07	5.31	7.69***	11.44	4.04	6.91***

Note. OR = odds ratio; SE = standard error.

* $p < .05$. *** $p < .001$.

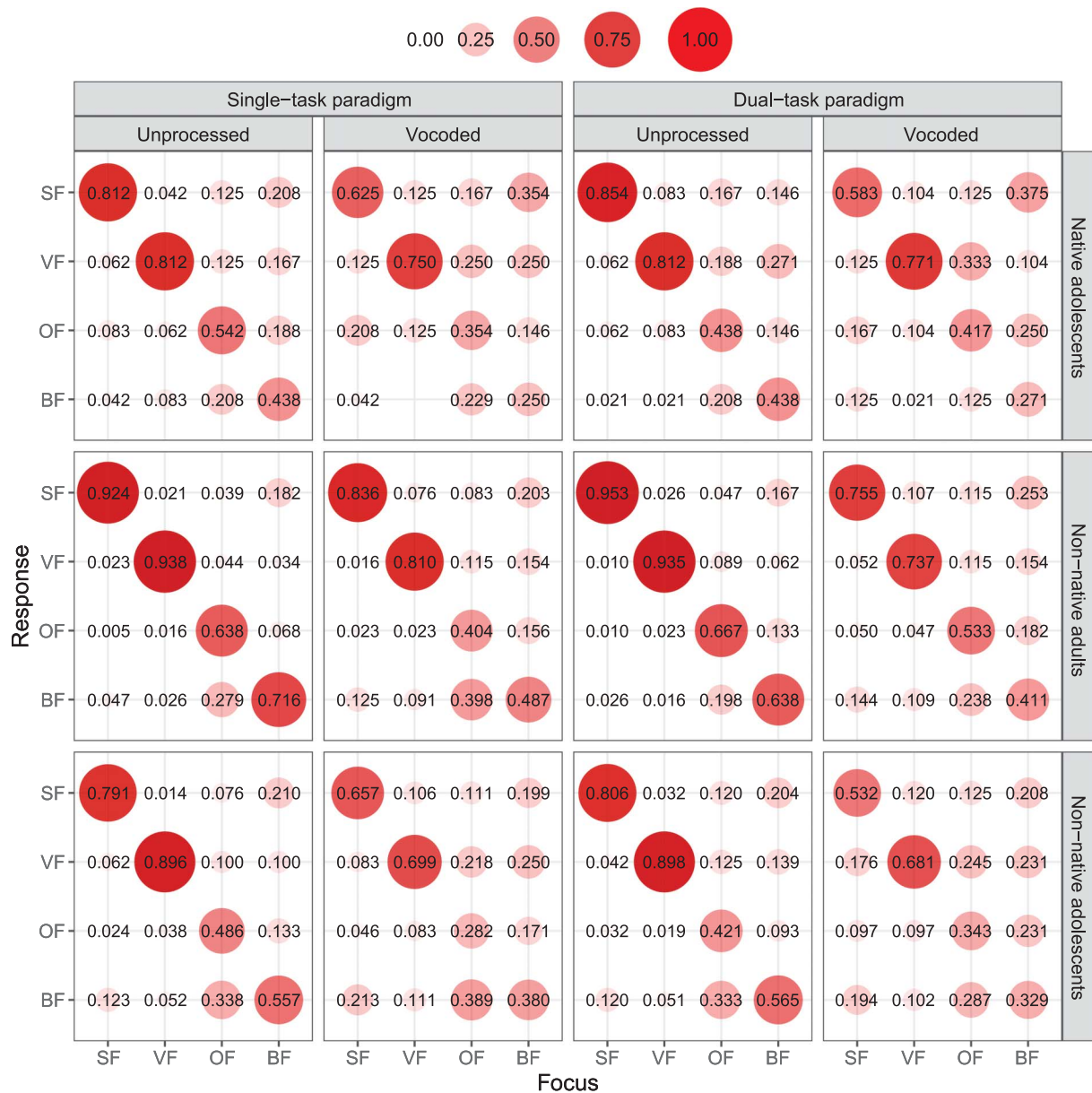
spectrotemporal degradations that can occur in electric hearing influence the ability of NH native Dutch learners of English with different proficiency levels to recognize the intended meaning of English utterances based on the position of the prosodically focused word as well as how effortful it is for these listeners to process prosodic focus patterns in the non-native language.

Overall, this study showed that, as predicted, the CI simulation has a significant impact on the interpretation of prosodically marked focus. Listeners were less accurate in identifying the correct context question in the CI-simulated condition compared with the non-CI-simulated condition. This is in line with previous research that has shown that CI users and NH listeners listening to CI-simulated speech are less accurate in recognizing prosodically focused words in a native language (Holt et al., 2016; Kalathottukaren et al., 2015; Meister et al., 2009, 2011; O’Halpin, 2010; van de Velde et al., 2017). In our study, both the native and non-native listeners were significantly impacted by the CI simulation. This indicates that, regardless of whether listeners are native or non-native listeners, the interpretation of prosodic focus is compromised

when prosodic cues in the speech signal are obscured due to the spectrotemporal degradation of the simulated electric speech signal (Başkent et al., 2016; Chatterjee & Peng, 2008; Gaudrain & Başkent, 2018; Pisoni, 2005; Shannon et al., 2004). This outcome reiterates the importance of quality access to prosodic cues for the ability to recognize the intended meaning of prosodic patterns (for an overview of similar findings in the literature, see Everhardt et al., 2020). This study thus adds to the body of literature showing a negative effect of CI processing on various aspects of prosody perception including linguistic focus, such as word stress (e.g., Lyxell et al., 2009; Morris et al., 2013), boundary marking (e.g., Kalathottukaren et al., 2015; Morris et al., 2013), questions versus statements (e.g., Kalathottukaren et al., 2015; Meister et al., 2009; van Zyl & Hanekom, 2013), or emotional prosody (e.g., Agrawal et al., 2012; Chatterjee et al., 2015; Kalathottukaren et al., 2015; Luo et al., 2007).

Furthermore, the results revealed that the interpretation of prosodically marked linguistic focus varies between different focus types. Listeners were consistently less accurate in linking the OF and BF stimuli to the OF and BF

Figure 3. Mean proportion of subject focus (SF), verb focus (VF), object focus (OF), and broad focus (BF) responses per focus type (SF, VF, OF, and BF) for each processing condition (unprocessed and vocoded), participant group (native adolescents, non-native adults, and non-native adolescents), and task paradigm (single-task and dual-task). Symbol size and color transparency reflect the proportion of responses.



context questions, respectively, compared with how well they could link the SF and VF stimuli to their respective context questions. As previously discussed, the nuclear accent is by default on the phrase-final word, whereas the accent is on the focused word for utterances with narrow focus (Calhoun, 2010; Cole, 2015; Ladd, 2008). The nuclear accent is thus on the object for both the OF and BF stimuli of this study. It is likely that the listeners experienced confusion between OF and BF stimuli, resulting in the reduced accuracy for these two focus types. Previous

research has also shown that listeners have a hard time differentiating between broad focus and narrow focus when the nuclear accent is in the same position for these two focus types (e.g., Gussenhoven, 1983; Roettger et al., 2019; Welby, 2003). The focus interpretation matrix of this study showed that when listeners heard OF stimuli and did not correctly link it to the OF context question, they most frequently identified the BF context question as the correct context question. Listeners thus regularly confused the narrow focus type with the broad focus type,

Table 5. Probability of a correctly recalled name for each recall trial (1–6).

Recall trial	Est.	SE	95% CI
Recall Trial 1	0.625	0.06	[0.51, 0.73]
Recall Trial 2	0.586	0.05	[0.48, 0.69]
Recall Trial 3	0.520	0.06	[0.41, 0.63]
Recall Trial 4	0.556	0.04	[0.48, 0.63]
Recall Trial 5	0.761	0.03	[0.69, 0.82]
Recall Trial 6	0.964	0.01	[0.94, 0.98]

Note. Est. = estimate; SE = standard error; CI = confidence interval.

whereas the opposite was not always the case. This confusion can be explained by the phonetic prominence of the relevant constituents in these stimuli. The acoustic analysis (described in the Method section above) showed that BF stimuli contain a prenuclear pitch accent on the subject with higher absolute phonetic prominence than that of the nuclear accent on the object. Listeners are able to recognize downstepped nuclear accents as structurally more prominent even if the preceding prenuclear accented constituent is phonetically more prominent (e.g., Ayers, 1996; Rump & Collier, 1996; Terken & Hermes, 2000), and listeners in this study did also frequently link BF stimuli to the BF context question. Yet, in those cases when the nuclear accent position did not determine the interpretation of BF stimuli, listeners mainly relied on the absolute phonetic prominence.

The results of the primary task also showed that the focus interpretation accuracy was higher for the non-native adults compared with the non-native adolescents,

Table 6. Differences in the probability of a correctly recalled name between the recall trials (1–6).

Recall trial comparison	OR	SE	z
Recall Trial 1 vs. 2	1.18	0.14	1.34
Recall Trial 1 vs. 3	1.54	0.18	3.59**
Recall Trial 1 vs. 4	1.33	0.22	1.75
Recall Trial 1 vs. 5	0.52	0.11	–3.02*
Recall Trial 1 vs. 6	0.06	0.02	–9.38***
Recall Trial 2 vs. 3	1.30	0.15	2.32
Recall Trial 2 vs. 4	1.13	0.16	0.84
Recall Trial 2 vs. 5	0.44	0.08	–4.34***
Recall Trial 2 vs. 6	0.05	0.01	–10.55***
Recall Trial 3 vs. 4	0.87	0.13	–0.95
Recall Trial 3 vs. 5	0.34	0.07	–5.41***
Recall Trial 3 vs. 6	0.04	0.01	–11.23***
Recall Trial 4 vs. 5	0.39	0.06	–6.63***
Recall Trial 4 vs. 6	0.05	0.01	–11.93***
Recall Trial 5 vs. 6	0.12	0.03	–8.83***

Note. OR = odds ratio; SE = standard error.

* $p < .05$. ** $p < .01$. *** $p < .001$.

though this was only significant for SF and OF stimuli. The difference in focus interpretation accuracy between these two groups could thus indicate that the accuracy of focus recognition increases with increasing proficiency in the non-native language, as was predicted based on previous research (e.g., Baker, 2010). However, given the small sample size of the native control group, it is hard to draw any conclusions from this finding, and we acknowledge this limitation to the study. Future research with larger sample sizes will have to show whether the differences previously found between native and non-native adults (Akker & Cutler, 2003; Baker, 2010; Wayland et al., 2019) can also be observed in adolescents or whether such differences are not yet prevalent during adolescence due to potentially longer developmental periods for prosody-related tasks (e.g., Cutler & Swinney, 1987; Nagels et al., 2020). As such, we acknowledge that it is not possible to disentangle non-native prosody perception abilities from overall prosody perception abilities in this study.

In terms of listening effort during prosodic focus processing, the results revealed no difference in the performance in the secondary task between the CI-simulated and non-CI-simulated stimuli. We expected that the secondary task would yield lower scores for the adverse CI-simulated condition, based on previous dual-task studies (Pals et al., 2013; Sarampalis et al., 2009), which would be indicative of an increase in listening effort for the primary focus interpretation task in this condition. Instead, the processing distinction was not a necessary predictor variable for the secondary task, implying that prosodic focus processing is similarly effortful in the CI-simulated and non-CI-simulated conditions. Similarly, there were no significant differences between the groups for the secondary task, indicating that listening effort was not increased for the non-native listeners even though this was predicted based on previous research (e.g., Borghini & Hazan, 2018, 2020). Furthermore, the lack of an interaction between the groups and processing conditions in the secondary task suggests that listening effort was not higher for the non-native listeners than for the native listeners in the CI-simulated condition, nor for the less proficient non-native listeners compared with the more proficient non-native listeners in the CI-simulated condition. These results may suggest that prosodic focus identification happens more automatically than we expected and, as such, does not draw extensive cognitive resources. It may also be the case, however, that the absence of a secondary-task effect is indicative of the memory task not being sensitive (or appropriate) enough to capture changes in the effortfulness of the primary task. The specific task parameters (such as the number of items to hold in memory) or the repeating nature of the sentences may have contributed to this. As such, we hesitate to say with confidence how the

effortfulness of prosody perception is affected by CI processing or language.

Finally, this study revealed no significant interaction between the groups and processing conditions for the primary task, indicating that the influence of the CI simulation on the focus interpretation patterns did not differ between listener groups. In previous research on non-prosodic listening skills, differences have been observed in the impact of adverse listening conditions between native and non-native listeners. On the one hand, the difference in phoneme and sentence recognition accuracy between native and non-native listeners was found to be less prominent in more adverse CI-simulated conditions than in less adverse conditions (Yang et al., 2022). On the other hand, speech-in-noise perception studies have indicated that non-native listeners are more strongly affected by adverse listening conditions than native listeners (for a review, see Lecumberri et al., 2010). In our study, the non-native listeners were less accurate in linking the focus pattern to the correct context question in the CI-simulated condition compared with the non-CI-simulated condition, in line with previous research showing a clear disadvantage for non-native CI users compared with their non-native NH peers (Beeres-Scheenstra et al., 2017, 2020). However, the non-native listeners did not show a more extensive decrease in perceptual accuracy in the CI-simulated condition than the native listeners. Moreover, the less proficient non-native listeners did not show a more extensive decrease than the more proficient non-native listeners. The spectrotemporally degraded electric speech signal thus influences all listener groups similarly, indicating that the interpretation of prosodically marked linguistic focus is impacted by a CI simulation regardless of the language background of the listener.

Conclusions

This study assessed how a CI simulation influences the interpretation of prosodically marked linguistic focus in a non-native language and how it influences listening effort during non-native focus processing for non-native listeners with different proficiency levels. The results confirmed that a CI simulation significantly impacts how well listeners can link a prosodic focus pattern to the correct context question, yet there was no increase in listening effort for the more adverse CI-simulated condition. Moreover, we found no compelling evidence in support of a difference in focus interpretation accuracy across focus types between the native and non-native listeners or between non-native listeners with different proficiency levels in the non-native language. We also found that listening effort was not increased for the non-native listeners compared

with native listeners or for the less proficient non-native listeners compared with the more proficient non-native listeners. Finally, the results revealed that the influence of the spectrotemporally degraded electric speech signal does not interact with listener group, indicating that non-native listeners are not more strongly affected by a CI simulation than native listeners and that the impact of the CI simulation is not modulated by age or proficiency level in the non-native language.

Data Availability Statement

The data and code for this study are available at <https://doi.org/10.34894/HCNIE8>.

Acknowledgments

This research was supported by the Center for Language and Cognition Groningen, the Netherlands Organisation for Scientific Research, the Netherlands Organisation for Health Research and Development (VICI 918-17-603, recipient: Deniz Başkent), and the Heinsius Houbolt Foundation.

References

- Agrawal, D., Timm, L., Viola, F. C., Debener, S., Büchner, A., Dengler, R., & Wittfoth, M. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. *BMC Neuroscience*, 13(1), Article 113. <https://doi.org/10.1186/1471-2202-13-113>
- Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, 6(2), 81–96. <https://doi.org/10.1017/S1366728903001056>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Ayers, G. M. (1996). *Nuclear accent types and prominence: Some psycholinguistic experiments* (Publication No. 304278671) [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations & Theses A&I.
- Baker, R. E. (2010). Non-native perception of native English prominence. *Proceedings of the 5th International Conference on Speech Prosody*, No. 171. <http://sprosig.org/sp2010/papers/100171.pdf>
- Başkent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Eds.), *Scientific foundations of audiology: Perspectives from physics, biology, modeling, and medicine* (pp. 285–319). Plural.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Beeres-Scheenstra, R. J., Azar, A. R., Heinzmann, S., Stieger, C., Kompis, M., Caversaccio, M., Bodmer, D., Huber, A., Lehnick, D., Candreia, C., & Linder, T. E. (2020). Multicenter study investigating foreign language acquisition at school in children, adolescents, and young adults with uni- or bilateral cochlear implants in the Swiss German population. *Otology & Neurotology*, 41(5), e580–e587. <https://doi.org/10.1097/MAO.0000000000002607>
- Beeres-Scheenstra, R., Ohnsorg, C., Candreia, C., Heinzmann, S., Castellanos, S., De Min, N., & Linder, T. E. (2017). Single-center study investigating foreign language acquisition at school in children, adolescents, and young adults with uni- or bilateral cochlear implants in the Swiss German population. *Otology & Neurotology*, 38(6), 833–838. <https://doi.org/10.1097/MAO.0000000000001431>
- Birch, S., & Clifton, C., Jr. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, 38(4), 365–391. <https://doi.org/10.1177/002383099503800403>
- Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 239–270). De Gruyter Mouton. <https://doi.org/10.1515/9783110261790.239>
- Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.14) [Computer software]. <https://www.praat.org>
- Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupilometry study. *Frontiers in Neuroscience*, 12, 152. <https://doi.org/10.3389/fnins.2018.00152>
- Borghini, G., & Hazan, V. (2020). Effects of acoustic and semantic cues on listening effort during native and non-native speech perception. *The Journal of the Acoustical Society of America*, 147(6), 3783–3794. <https://doi.org/10.1121/10.0001126>
- Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press. <https://doi.org/10.1037/10037-000>
- Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86(1), 1–42. <https://doi.org/10.1353/lan.0.0197>
- Calhoun, S., Wollum, E., & Kruse Va'ai, E. (2021). Prosodic prominence and focus: Expectation affects interpretation in Samoan and English. *Language and Speech*, 64(2), 346–380. <https://doi.org/10.1177/0023830919890362>
- Chatterjee, M., & Peng, S.-C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235(1–2), 143–156. <https://doi.org/10.1016/j.heares.2007.11.004>
- Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., & Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*, 322, 151–162. <https://doi.org/10.1016/j.heares.2014.10.003>
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1–2), 1–31. <https://doi.org/10.1080/23273798.2014.963130>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <http://www.coe.int/lang-cefr>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2), 141–201. <https://doi.org/10.1177/002383099704000203>
- Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, 14(1), 145–167. <https://doi.org/10.1017/S0305000900012782>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/S13428-014-0458-Y>
- Everhardt, M. K., Sarampalis, A., Coler, M., Başkent, D., & Lowie, W. (2019). Perception of L2 lexical stress in words degraded by a cochlear implant simulation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 102–106). Australasian Speech Science and Technology Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_151.pdf
- Everhardt, M. K., Sarampalis, A., Coler, M., Başkent, D., & Lowie, W. (2020). Meta-analysis on the identification of linguistic and emotional prosody in cochlear implant users and vocoder simulations. *Ear and Hearing*, 41(5), 1092–1102. <https://doi.org/10.1097/AUD.0000000000000863>
- Everhardt, M. K., Sarampalis, A., Coler, M., Başkent, D., & Lowie, W. (2022). Interpretation of prosodically marked focus in cochlear implant-simulated speech by non-native listeners. *Proceedings of the 11th International Conference on Speech Prosody*, 72–76. <https://doi.org/10.21437/SpeechProsody.2022-15>
- Flège, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470. <https://doi.org/10.1006/jpho.1997.0052>
- Gaudrain, E. (2016). *Vocoder* (Version 1.0) [MATLAB code]. GitHub. <https://github.com/egaudrain/vocoder>
- Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, 39(2), 226–237. <https://doi.org/10.1097/AUD.0000000000000480>
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 351–360. [https://doi.org/10.1016/S0022-5371\(66\)80044-0](https://doi.org/10.1016/S0022-5371(66)80044-0)
- Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech*, 26(1), 61–80. <https://doi.org/10.1177/002383098302600104>
- Holt, C. M., Demuth, K., & Yuen, I. (2016). The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants. *Ear and Hearing*, 37(4), e256–e262. <https://doi.org/10.1097/AUD.0000000000000253>
- Inspectie van het Onderwijs. (2019). *Peil.Engels—Technische rapportage* [Peil.Engels—Technical report]. <https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2019/11/08/technisch-rapport-peil.engels/Peil.Engels+TECHNISCHE+RAPPORTAGE.pdf>
- Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.
- Kalathottukaren, R. T., Purdy, S. C., & Ballard, E. (2015). Prosody perception and musical pitch discrimination in adults using cochlear implants. *International Journal of Audiology*, 54(7), 444–452. <https://doi.org/10.3109/14992027.2014.997314>
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Lange, K., Kühn, S., & Filevich, E. (2015). “Just Another Tool for Online Studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, 10(7), Article e0130834. <https://doi.org/10.1371/journal.pone.0130834>
- Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, 52(11–12), 864–886. <https://doi.org/10.1016/j.specom.2010.08.014>
- Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.7.5) [R package]. CRAN. <https://CRAN.R-project.org/package=emmeans>

- Luo, X., Fu, Q.-J., & Galvin, J. J., III. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, 11(4), 301–315. <https://doi.org/10.1177/1084713807305301>
- Lyxell, B., Wass, M., Sahlén, B., Samuelsson, C., Asker-Árnason, L., Ibertsson, T., Mäki-Torkko, E., Larsby, B., & Hällgren, M. (2009). Cognitive development, reading and prosodic skills in children with cochlear implants. *Scandinavian Journal of Psychology*, 50(5), 463–474. <https://doi.org/10.1111/j.1467-9450.2009.00754.x>
- Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M. (2011). The perception of sentence stress in cochlear implant recipients. *Ear and Hearing*, 32(4), 459–467. <https://doi.org/10.1097/AUD.0b013e3182064882>
- Meister, H., Landwehr, M., Pyschny, V., Walger, M., & von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, 48(1), 38–48. <https://doi.org/10.1080/14992020802293539>
- Moberly, A. C., Lewis, J. H., Vasil, K. J., Ray, C., & Tamati, T. N. (2021). Bottom-up signal quality impacts the role of top-down cognitive-linguistic processing during speech recognition by adults with cochlear implants. *Otology & Neurotology*, 42(10S), S33–S41. <https://doi.org/10.1097/mao.0000000000003377>
- Morris, D., Magnusson, L., Faulkner, A., Jönsson, R., & Juul, H. (2013). Identification of vowel length, word stress, and compound words and phrases by postlingually deafened cochlear implant listeners. *Journal of the American Academy of Audiology*, 24(9), 879–890. <https://doi.org/10.3766/jaaa.24.9.11>
- Nagels, L., Gaudrain, E., Vickers, D., Lopes, M. M., Hendriks, P., & Başkent, D. (2020). Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ*, 8, Article e8773. <https://doi.org/10.7717/peerj.8773>
- O’Halpin, R. (2010). *The perception and production of stress and intonation by children with cochlear implants* [Doctoral dissertation, University College London]. UCL Discovery. <https://discovery.ucl.ac.uk/id/eprint/20406>
- Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research*, 56(4), 1075–1084. [https://doi.org/10.1044/1092-4388\(2012/12-0074\)](https://doi.org/10.1044/1092-4388(2012/12-0074))
- Peng, Z. E., & Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. *Journal of Speech, Language, and Hearing Research*, 62(4), 1068–1081. https://doi.org/10.1044/2018_JSLHR-H-17-0423
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37(1), 5S–27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Pisoni, D. B. (2005). Speech perception in deaf children with cochlear implants. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 494–523). Blackwell Publishing. <https://doi.org/10.1002/9780470757024.ch20>
- Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning—The case of information structure in American English. *Language, Cognition and Neuroscience*, 34(7), 841–860. <https://doi.org/10.1080/23273798.2019.1587482>
- Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, 39(1), 1–17. <https://doi.org/10.1177/002383099603900101>
- Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, 52(5), 1230–1240. [https://doi.org/10.1044/1092-4388\(2009/08-0111\)](https://doi.org/10.1044/1092-4388(2009/08-0111))
- Shannon, R. V., Fu, Q.-J., Galvin, J. J., III, & Friesen, L. (2004). Speech perception with cochlear implants. In F.-G. Zeng, A. N. Popper, & R. R. Fay (Eds.), *Cochlear implants: Auditory prostheses and electric hearing* (pp. 334–376). Springer. https://doi.org/10.1007/978-0-387-22585-2_8
- Smits, C., Theo Goverts, S., & Festen, J. M. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America*, 133(3), 1693–1706. <https://doi.org/10.1121/1.4789933>
- Smits, C., Watson, C. S., Kidd, G. R., Moore, D. R., & Theo Goverts, S. (2016). A comparison between the Dutch and American-English digits-in-noise (DIN) tests in normal-hearing listeners. *International Journal of Audiology*, 55(6), 358–365. <https://doi.org/10.3109/14992027.2015.1137362>
- Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 89–127). Kluwer. https://doi.org/10.1007/978-94-015-9413-4_5
- Vallduví, E. (2016). Information structure. In M. Aloni & P. Dekker (Eds.), *The Cambridge handbook of formal semantics* (pp. 728–755). Cambridge University Press. <https://doi.org/10.1017/CBO9781139236157.024>
- van de Velde, D. J., Schiller, N. O., van Heuven, V. J., Levelt, C. C., van Ginkel, J., Beers, M., Briaire, J. J., & Frijns, J. H. M. (2017). 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*, 141(5), 3349–3363. <https://doi.org/10.1121/1.4982198>
- 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. <https://doi.org/10.1016/j.jcomdis.2013.09.002>
- Wayland, R., Guerra, C., Chen, S., & Zhu, Y. (2019). English focus perception by Mandarin listeners. *Language*, 4(4), Article 91. <https://doi.org/10.3390/languages4040091>
- Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, 46(1), 53–81. <https://doi.org/10.1177/00238309030460010401>
- Wells, B., Peppé, S., & Goulondris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31(4), 749–778. <https://doi.org/10.1017/S030500090400652X>
- Yang, J., Wagner, A., Zhang, Y., & Xu, L. (2022). Recognition of vocoded speech in English by Mandarin-speaking English-learners. *Speech Communication*, 136, 63–75. <https://doi.org/10.1016/j.specom.2021.11.008>