University of Groningen

The impact of speaking style on speech recognition in quiet and multi-talker babble in adult cochlear implant users

Rodman, Cole; Moberly, Aaron C.; Janse, Esther; Baskent, Deniz; Tamati, Terrin N.

Published in:
Journal of the Acoustical Society of America

DOI:
10.1121/1.5141370

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:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Rodman, C., Moberly, A. C., Janse, E., Baskent, D., & Tamati, T. N. (2020). The impact of speaking style on speech recognition in quiet and multi-talker babble in adult cochlear implant users. *Journal of the Acoustical Society of America*, *147*(1), 101-107. https://doi.org/10.1121/1.5141370

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.

The impact of speaking style on speech recognition in quiet and multi-talker babble in adult cochlear implant users

Cole Rodman, Aaron C. Moberly, Esther Janse, Deniz Başkent, and Terrin N. Tamati

Citation: The Journal of the Acoustical Society of America **147**, 101 (2020); doi: 10.1121/1.5141370
View online: https://doi.org/10.1121/1.5141370
View Table of Contents: https://asa.scitation.org/toc/jas/147/1
Published by the Acoustical Society of America

The impact of speaking style on speech recognition in quiet and multi-talker babble in adult cochlear implant users

Cole Rodman,[1] Aaron C. Moberly,[1] Esther Janse,[2] Deniz Başkent,[3] and Terrin N. Tamati[1,a)]
[1]*Department of Otolaryngology—Head and Neck Surgery, The Ohio State University Wexner Medical Center, 915 Olentangy River Road, Suite 4000, Columbus, Ohio 43212, USA*
[2]*Centre for Language Studies, Radboud University Nijmegen, Nijmegen, The Netherlands*
[3]*Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*

(Received 30 September 2019; revised 12 November 2019; accepted 30 November 2019; published online 17 January 2020)

The current study examined sentence recognition across speaking styles (conversational, neutral, and clear) in quiet and multi-talker babble (MTB) for cochlear implant (CI) users and normal-hearing listeners under CI simulations. Listeners demonstrated poorer recognition accuracy in MTB than in quiet, but were relatively more accurate with clear speech overall. Within CI users, higher-performing participants were also more accurate in MTB when listening to clear speech. Lower performing users' accuracy was not impacted by speaking style. Clear speech may facilitate recognition in MTB for high-performing users, who may be better able to take advantage of clear speech cues.

© *2020 Acoustical Society of America*. https://doi.org/10.1121/1.5141370

[DDOS] Pages: 101–107

I. INTRODUCTION

For individuals with cochlear implants (CIs), understanding speech in real-world conditions can be incredibly difficult. CI users rely upon a speech signal that is spectro-temporally degraded due to limitations in information transmission of electric stimulation of the auditory nerve (Başkent *et al.*, 2016b). In real-world listening environments, further signal degradation comes from environmental conditions (e.g., noise, masking speech), and the acoustic-phonetic variability from across talkers (e.g., gender, age, regional or foreign accent) and within talkers (e.g., speaking style, emotion) (Mattys *et al.*, 2012; Gilbert *et al.*, 2013). Understanding speech in the presence of competing talkers (or "babble"), conversational speech with reduced speech cues (Liu *et al.*, 2004; Tamati *et al.*, 2019), and high talker variability (Faulkner *et al.*, 2015), have all been shown to be challenging for CI users.

In real-world conditions, speakers may improve the clarity of their speech by speaking more loudly, slowing their speech, or hyperarticulating (Krause and Braida, 2002, 2004; Hazan *et al.*, 2018). In normal hearing (NH) listeners, these "clear speech" modifications typically result in an intelligibility benefit (Janse *et al.*, 2007; Liu *et al.*, 2004) relative to conversationally reduced speech, where speech sounds are often shorter or weaker, while the speaking rate is often faster and more variable (e.g., Ernestus and Warner, 2011). In quiet, NH listeners are typically able to understand conversational reduced speech (Ernestus and Warner, 2011), although this comes at the cost of increased cognitive effort (Van Engen *et al.*, 2012). In background noise or competing talkers (Schum, 1996; Helfer, 1997), or when listening with a hearing impairment (Janse and Ernestus, 2011), however, listeners show a relatively greater benefit of clear speech over conversational reduced speech. Similarly, CI users have previously been shown to benefit from clear speech in quiet and in steady-state noise conditions, with greater overall benefit in noise (Iverson and Bradlow, 2002; Liu *et al.*, 2004; Smiljanic and Sladen, 2013), although potentially to a lesser degree than NH listeners (Smiljanic and Sladen, 2013). Thus, speaking style interacts with presentation conditions, such that clear speech results in a relatively greater benefit to accurate speech understanding in adverse listening conditions (noise) compared to more favorable conditions (quiet).

While previous research has suggested that CI users broadly benefit from clear speech in quiet and in noise, it is unclear if CI users show a similar benefit in the presence of multi-talker babble (MTB) and how that might vary by individual listener. Speech recognition with competing sound sources is considered one of the largest limitations for CI users (for a review, see Başkent *et al.*, 2016b). Compared to relatively simple noise competitors, more ecologically valid maskers, such as MTB, result in even larger differences in speech recognition accuracy between CI users and NH listeners (e.g., Friesen *et al.*, 2001; Stickney *et al.*, 2004). Previous findings suggest that CI users are unable to detect acoustic differences between the target and masking speech, such as voice cue differences, and are thereby impaired in using these cues to engage perceptual or linguistic mechanisms to segregate the target from the masking speech (e.g., Luo *et al.*, 2009; Gaudrain *et al.*, 2007, 2008; El Boghdady *et al.*, 2019). In NH listeners, it has been widely demonstrated that effective segregation of the target from masking speech also depends on several linguistic factors, including speaking style, as well as the linguistic content of the target and the masking speech (e.g., Calandruccio *et al.*, 2010, 2014). Further, the benefit of a clear speaking style has been found to vary by the masker type and signal-to-noise ratio

---

[a)] Electronic mail: terrin.tamati@osumc.edu

(SNR) (e.g., Van Engen *et al.*, 2014; Calandruccio *et al.*, 2010). Since speech in MTB is limited by reduced spectral resolution in CI users, the effect of speaking style may also differ across listening conditions in CI users. Further, differences in signal quality as well as the linguistic or cognitive skills of the listener may contribute to individual differences in speech recognition in MTB.

Thus, the first aim of the current study was to determine the effect of speaking style and background competition on speech recognition in CI users and simulated CIs (8- or 4-channel acoustic simulations of CI hearing, to cover a wide range of performance, e.g., Friesen *et al.*, 2001). To examine the interaction of speaking style and background competition, we compared word-in-sentence recognition accuracy across three distinct speaking styles, read text (clear speech), retold stories (neutral speech), and conversational reduced (conversational speech; following Tamati *et al.*, 2018) in quiet and in 4-talker MTB. Previous findings imply that CI users would benefit from clear speech in quiet and in the presence of MTB, with a relatively greater benefit for MTB. Alternatively, limitations in CI hearing, associated with a deficit in discriminating speaking style differences (Tamati *et al.*, 2019), may reduce the benefit afforded by a clear speaking style; that is, MTB may actually further limit access to relevant clear speech cues, resulting in a lack of benefit of clear over conversational speech in MTB. Given the vast individual differences attested in CI users (Lazard *et al.*, 2012; Blamey *et al.*, 2013), the second aim of the study was to investigate whether individual differences in speech recognition determine the extent to which CI users are able to benefit from clear speech modifications in quiet and in MTB. Liu *et al.* (2004) demonstrated that only the higher-performing CI users showed an advantage for clear speech over conversational speech in steady-state noise, while both lower- and higher-performing CI users benefited from clear speech over conversational speech in quiet. Thus, the overall goal of the current study was to explore the relationships between speaker style, MTB, and speech understanding in individual CI users.

II. METHODS

A. Listeners

Ten native Dutch speaking, experienced CI users [age 38–75 years; $M = 68$, standard deviation (SD) = 11.3; 3 female] participated in the study (see Table I for demographics). All had used their implants for at least 2.5 years (2.5–13 years) and were implanted after age 18 years.

Twenty young, native, NH Dutch speakers (age 20–29 years; $M = 20.6$; SD = 1.5; 15 female; 25 dB hearing level or better at audiometric frequencies 250–8000 Hz) participated in the current study. Participants were randomly divided into two groups: the 8-channel (CI-8) and 4-channel (CI-4) CI-simulation conditions.

All participants received a detailed explanation of the study and signed an informed written consent. For NH listeners, compensation was 8 euros or partial course credit for 1 h of testing. For CI users, compensation was 16 euros for participating in a larger study, which included the current set of experiments and lasted approximately 2 h in total. The study was approved by the Medical Ethics Committee of the UMCG (METc2012-455).

B. Materials

Materials consisted of 72 sentence-length utterances produced by two talkers (1 female/1 male) selected from the Instituut voor Fonetische Wetenschappen Amsterdam corpus of the Institute of Phonetic Sciences Amsterdam (van Son *et al.*, 2001). For each talker, 12 utterances were produced each in the context of a conversation (conversational reduced—“conversational”), from the retelling of a story (retold story—“neutral”), and from a read list (read text—“clear”), for 36 in total. A full description of the acoustic-phonetic characteristics of a larger set of materials from which the stimuli were selected can be found in the study methods provided in Tamati *et al.* (2019). As summarized in Tamati *et al.* (2019), the clear speech (read text) originating from the larger corpus demonstrated properties consistent with a carefully articulated speaking style: a greater relative number of pauses, a slower speaking rate (although varying across talkers), a higher average $F0$ and $F0$ range, and more fully realized sound segments, including more frequent word-final [t]-realization, schwa realization in unstressed syllables, word-final [n]-realization, and postvocalic–[r]realization. The characteristics of the clear speech are described in contrast with the conversational speech originating from the larger corpus (conversational reduced), which demonstrated features more consistent with conversational speech: faster speaking rate, a lower average $F0$ and $F0$ range, and more frequent

TABLE I. Demographic information of CI users.

Participant	Age (years)	Gender	Etiology	Age at Onset of Hearing Loss (years)	Duration CI Use (years)	Device	Configuration
CI1	67	M	Genetic—progressive	13	3	Advanced Bionics	CI L
CI2	75	M	Traumatic Head Injury	68	8	Cochlear	CI R
CI3	78	F	Unknown	0	10	Cochlear	CI R
CI4	68	M	Autoimmune	29	10	Cochlear	CI L
CI5	75	M	Genetic—progressive	50	9	Advanced Bionics	Bilateral
CI6	68	M	Viral—sudden	61	6	Cochlear	CI R
CI7	66	F	Unknown—progressive	34	2.5	Advanced Bionics	CI R
CI8	38	M	Genetic—progressive	1	13	Cochlear	CI R, HA L
CI9	70	M	Unknown	55	3	Advanced Bionics	CI R, HA L
CI10	60	F	Genetic—progressive	17	13	Cochlear	CI R, HA L

reduction/deletion of the four sound segments. The neutral speech (retold story) displayed properties of both clear and conversational speech and presented an in-between case for some measures: slower speaking rate, fairly high average $F0$ but decreased $F0$ range, frequent deletion of word-final [t], moderate schwa realization in unstressed syllables, moderate deletion of word-final [n], and frequent realization of post-vocalic. The features of these three speech categories are largely consistent with previous descriptions of speaking style differences among scripted speech and variations of non-scripted speech in Dutch (Ernestus *et al.*, 2015).

C. Prodedure

Participants were tested individually, seated in an anechoic room. Stimulus materials were equal in intensity and presented at 65 dB sound pressure level (SPL), via a loudspeaker (Precision 80, Tannoy, Coatbridge, United Kingdom) placed approximately 1 m from the participant at 0° azimuth. For the experiment, CI participants used their everyday CI settings set to a comfortable volume.

Half of the sentences were presented in quiet (block 1) and half were presented in MTB at +10 dB SNR (block 2). The block order (quiet-MTB) was the same for all participants. Each block contained 6 conversational, 6 neutral, and 6 clear sentences, presented in random order and only once without repetition. For the MTB condition, the target sentences were mixed with random samples of four-talker babble made from samples of conversational speech produced by 2 male talkers and 2 female talkers (IFADV Corpus; van Son *et al.*, 2008).

On each trial, participants were presented with a single sentence and were asked to verbally repeat the words that they heard. Partial answers and guessing were encouraged. The participants' responses were recorded and scored offline by a native Dutch speaker. Exact word order was not required, but plural or possessive morphological markers were required to match the word.

For CI simulation, all stimuli were processed through an 8-channel (CI-8 listener group) or 4-channel (CI-4 listener group) noise-band vocoder with MATLAB code maintained by the dB SPL lab at the UMCG (e.g., Gaudrain and Başkent, 2015). The sentences (with or without MTB) were filtered into 4 or 8 frequency bands between 150 and 7000 Hz, using 12th order, zero-phase Butterworth filters. Greenwood's frequency-to-place mapping function was used such that each band corresponded to evenly spaced regions of the cochlea (Greenwood, 1990). Noise-band carriers were generated by filtering white noise into spectral bands using the same 12th order Butterworth bandpass filters. The stimuli were constructed by modulating the noise carriers in each channel with the corresponding extracted envelope, and adding together the modulated noise bands from all vocoder channels.

III. RESULTS

Recognition accuracy, as determined by the total number of words correctly identified, was measured for all three listener groups across speaking styles and background noise conditions. Overall (Fig. 1), mean accuracy for clear speech was highest ($M = 34.95\%$, SD = 38.20), followed by neutral speech ($M = 29.85\%$, SD = 35.46), and then conversational speech ($M = 27.90$, SD = 34.47). Mean accuracy for the Quiet condition ($M = 50.16\%$, SD = 37.22) was higher than for the MTB condition ($M = 11.64\%$, SD = 22.18). Recognition accuracy was highest in the CI-8 group ($M = 47.6\%$, SD = 38.6), lowest in the CI-4 group ($M = 18.3\%$, SD = 26.6), with the CI group in the middle ($M = 29.1\%$, SD = 35.1).

In order to examine the effects of speaking style and noise conditions on recognition accuracy across the three listener groups, a mixed effects model was created treating speaking style, noise condition, and listener group as fixed effects, participant as a random effect, and overall performance on the speaking style sentence recognition test—as measured in rational arcsine units (RAUs) (Studebaker, 1985)—as the outcome variable in R statistic software (Version 3.6.0, macOS Mojave version 10.14.4). Note that the intercept and the slopes of the noise condition and speaking style variables were all allowed to vary with the random variable, participant, as recent work has shown that inclusion of the maximal rational model structure in the random effect term yields more robust results (Barr *et al.*, 2013). Likelihood ratio (LR) testing was utilized to determine variables and model structure. The maximal model was created with interactions between all three variables (i.e., speaking style, noise condition, and listener group). LR testing for an interaction of speaking style and listener group [χ^2 (10) = 11.93, $p = 0.29$] or speaking style and noise condition [χ^2 (2) = 11.93, $p = 0.66$] did not prove significant, while LR testing for an interaction of listener group and noise condition [χ^2 (8) = 47.49, $p < 0.001$ [did prove significant. Main effects were significant for speaking style [χ^2 (2) = 19.51, $p < 0.001$] and noise condition [χ^2 (1) = 66.13, $p < 0.001$] and marginally significant for listener group [χ^2 (2) = 5.70, $p = 0.058$]. Thus, the final model included a linear

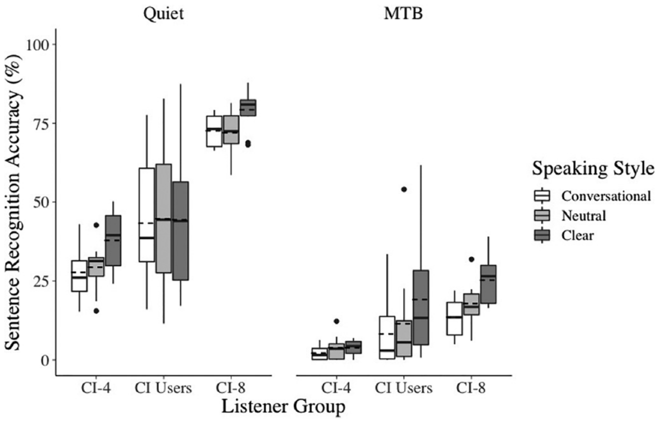

FIG. 1. Mean word-in-sentence recognition accuracy by listener group (CI-4, CI Users, and CI-8 users) and Speaking Style (Conversational, Neutral, Clear Speech) for Quiet and MTB noise conditions. The boxes extend from the lower to the upper quartile (the interquartile range, IQ), the solid midline indicates the median, and the dashed midline indicates the mean. The whiskers indicate the highest and lowest values no greater than 1.5 times the IQ, and the dots indicate the outliers, which are defined as data points larger than 1.5 times the IQ.

TABLE II. Results of mixed effects modeling of main effects. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Predictor	Level	Coefficient	Error	df	*p*-value	
Intercept		6.03	3.12	33.19	0.062	
Noise Condition	MTB	*ref*				
	Quiet	34.54	2.48	31.26	<0.001	***
Listener Group	CI	*ref*				
	CI-4	−10.33	4.39	32.00	0.025	*
	CI-8	6.88	4.39	32.00	0.127	
Speaking Style	Conversational	*ref*				
	Neutral	1.51	1.52	177.06	0.322	
	Clear	7.07	1.57	109.83	0.000	***
Interactions	Quiet and CI-4	−2.22	3.48	31.66	0.528	
	Quiet and CI-8	24.88	3.48	31.66	<0.001	***

combination of the three fixed effects as well as an interaction term between listener group and noise condition. The full results of the model can be found in Table II.

The main effect for noise condition had a positive coefficient ($b = 34.54$, $p < 0.001$) with MTB as the baseline, matching the observation that listeners were more accurate in the Quiet condition than the MTB condition. The main effect for the listener group had CI users as the baseline and coefficients for CI-4 ($b = -10.32$, $p = 0.025$) and CI-8 ($b = 6.88$, $p = 0.127$) demonstrated that CI-4 users did worse, on average, than CI and CI-8 users, who performed similarly. Finally, the main effect for speaking style used conversational speech as the baseline condition and coefficients for neutral speech ($b = 1.51$, $p = 0.322$), and clear speech ($b = 7.07$, $p < 0.001$) indicated that accuracy improved from worst to best, in that order. The interaction coefficients— with the MTB condition and the CI user group as the baseline—show that the amount of release from MTB noise masking (i.e., Quiet relative to Noise performance) was similar for the CI user and the CI-4 groups ($b = -2.22$, $p = 0.527$), but was larger for the CI-8 group ($b = 24.88$, $p < 0.001$).

Although the interaction of speaking style and noise condition by listener group was not significant, these factors may interact at an individual level, given the vast individual differences in performance within groups (see Fig. 1). Therefore, to further examine the relationship between speaking style and noise condition, a mixed effects model was utilized with performance in the MTB condition (in RAUs) as the outcome, individual performance in the Quiet condition (in RAU) and speaking style as fixed effects, as well as their interaction, and participant as a random effect. Using LR testing to compare different models, including an interaction between individual Quiet condition sentence recognition and speaking style was found to significantly improve model fit [$\chi^2(2) = 12.67$, $p < 0.001$]. Across speaking styles, better performance in the Quiet condition predicted better performance in the MTB condition ($b = 0.21$, $p < 0.001$). A significant interaction was found between Quiet condition sentence recognition and the clear speaking style ($b = 0.21$, $p < 0.001$) such that the association between performance levels in the two noise conditions is stronger in the clear speech condition than in the conversational speech condition. The full results of the model can be found in Table III. The relationship between performance levels can be seen in the slopes displayed in Fig. 2.

IV. DISCUSSION

A clear, rather than conversational, speaking style may be one means of improving speech recognition for CI users (Liu *et al.*, 2004; Tamati *et al.*, 2019), but the extent to which a clear speaking style may benefit listeners in MTB and other adverse listening environments is still unknown. The current study examined the interaction between speaking style and noise (quiet, MTB) on sentence recognition in CI users and NH listeners under CI simulation.

Listener group (CI users, CI-4, CI-8), noise condition (Quiet, MTB), and speaking style (clear speech, neutral speech, conversational speech) were found to significantly affect sentence recognition accuracy. CI users varied greatly in the overall sentence recognition accuracy, with CI-4 and CI-8 approximating the range of performance among the CI users. The most striking effect was that MTB resulted in drastic declines in performance across all speaking styles and

TABLE III. Results of mixed effects modeling of individual differences. ** $p<0.01$.

Predictor	Level	Coefficient	Error	df	*p*-value	
Intercept		−2.02	3.33	56.63	0.547	
Individual Performance in Quiet		0.21	0.06	61.02	0.001	**
Speaking Style	Conversational	*ref*				
	Neutral	−1.28	3.10	48.98	0.680	
	Clear	−4.60	3.26	49.80	0.164	
Interaction	Quiet and Neutral	0.09	0.06	49.05	0.147	
	Quiet and Clear	0.21	0.06	49.19	0.001	**

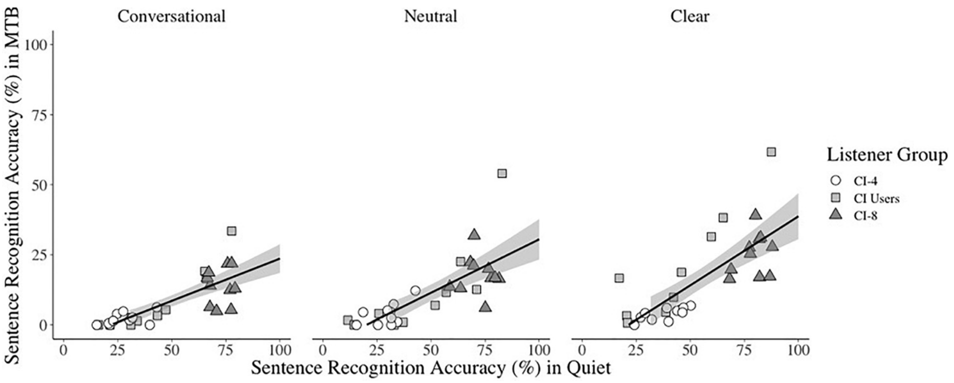

FIG. 2. Mean percent sentence recognition for the MTB noise condition (y axis) plotted against the mean percent sentence recognition in the Quiet condition (x axis) for Conversational Speech, Neutral Speech, and Clear Speech Speaking Styles, and for CI-4 users (circle), CI users (square), and CI-8 users (triangle). Linear regressions with 95% confidence intervals have also been plotted.

listener groups. Although consistent with our predictions, the magnitude of the effect of MTB on speech recognition is notable. For all participants, but especially the CI-4 listeners, accuracy scores were near floor at +10 dB SNR, suggesting that, in addition to the MTB, the task and materials might be quite challenging for CI users, perhaps due to the interleaved speaking style and talker variability and lack of strong semantic information with which listeners might compensate for the degraded conditions (Başkent *et al.*, 2016a, 2016b; see Tamati *et al.*, 2019 for additional information about the materials).

Across listeners, as expected, the CI-8 listeners were found to have the best performance across all tasks, while the CI-4 listeners had the poorest performance, and CI users were spread relatively evenly across the range of scores, confirming our design choice for approximating good and poor CI listening with 8- and 4-channel noise-vocoder simulations. A significant interaction was found between the listener group and noise condition, but no interaction between listener group and speaking style. CI-8 users were found to be disproportionately better under the MTB condition than either the CI or CI-4 users, consistent with previous research (Dorman *et al.*, 1998) and supporting the idea that increased spectral resolution likely provides the listeners with additional acoustic-phonetic details that can help in recognizing words in quiet and extracting linguistic content from words in a MTB background.

With regards to speaking style, consistent with previous findings (Liu *et al.*, 2004; Tamati *et al.*, 2019), CI users demonstrated worse performance with the conversational speech and better performance with the clear speech in both Quiet and MTB conditions, with neutral speech falling in the middle. These results support previous research demonstrating that CI users may benefit from clear speech relative to conversational reduced speech, which presents an additional cognitive and perceptual challenge (Liu *et al.*, 2004; Tamati *et al.*, 2019). However, the clear speech benefit was not affected by noise condition, with similar benefits broadly observed in both noise conditions and across listener groups. Iverson and Bradlow (2002) observed a benefit from clear speech on sentence recognition in speech-spectrum shaped noise conditions, with listeners demonstrating an even greater performance benefit from clear speech in noise. However, in these studies, speech understanding was near ceiling in quiet and much more accurate in comparable noise conditions (+10 dB SNR) to this study. Considering these previous findings, the current results again suggest that the noise condition may interact with speech materials and/or task demand, resulting in a poorer overall performance in MTB with the more difficult materials and task from the current study. As such, the clear speech benefit in noise or MTB may crucially depend on the range of performance, potentially resulting in a less clear speech benefit with overall performance closer to the ceiling or floor (see also Iverson and Bradlow, 2002). In the current study, the MTB condition was very challenging, with many participants near floor performance. As such, the MTB at this SNR may have obscured the speech cues too greatly, especially for lower-performing CI users, potentially impeding their ability to utilize clear speech cues to facilitate speech recognition. Future studies could use a range of SNRs and vary the number of talkers in the masker to obtain a larger range of performance in MTB and systematically explore possible interactions with the clear speech benefit and performance level.

Regarding individual differences, while a stronger clear speech benefit was not observed in MTB across groups, further analysis indicated that higher-performing CI users may have been better able to effectively utilize some clear speech cues to support speech recognition. Individuals who were most accurate in quiet conditions were performing disproportionately better in MTB when hearing clear speech, similar to findings from Liu *et al.* (2004). Similarly, there is evidence that some higher-performing CI users are better able to apply top-down compensatory strategies to improve recognition in adverse listening conditions (Bhargava *et al.*, 2014). These CI users may be able to better use predictive coding and downstream cognitive processing resources to free up resources to dedicate to the encoding of fine-grained acoustic details, potentially allowing them to take advantage of clear speech cues or engage in other compensatory strategies (Başkent *et al.*, 2016a; Moberly *et al.*, 2014, 2016).

Potential weaknesses of the current study should be noted. First, in the current study, sample sizes were relatively small with only ten participants per listener group. Additionally, the ten CI users varied greatly in age, age of implantation, device use, and likely language background and cognitive skills, which may influence sentence recognition accuracy (e.g., Schoof and Rosen, 2014). While the current study explored individual differences in the CI users' ability to benefit from clear speech modifications in quiet and in MTB, accounting for how these factors may contribute to the observed individual differences was beyond the scope of the current study. Additionally, the demographic characteristics of the CI users were not matched in the NH listener groups, hindering our ability to understand and account for group differences in the current study. Although CI users' performance was distributed relatively equally between the CI-4 and CI-8 listener groups—suggesting a similar effect of MTB across groups—differences in demographic characteristics, specifically age, may lead to different underlying processing strategies across speaking styles and MTB (e.g., Bhargava *et al.*, 2016). Therefore, larger studies involving more CI users and more carefully controlling for demographic characteristics and device use among the participants are needed to confirm the effect of speaking style and MTB in CI users and to explore the factors underlying individual differences.

The current study has provided a first step in understanding the interactions of speaking style and background noise, specifically in adult, post-lingually deafened CI users and how these interactions may vary depending on the individual CI user. Taken together, the results of this study demonstrate that CI users and NH listeners under CI simulation show poor speech recognition in the presence of MTB, but that clearer speaking styles can significantly improve sentence recognition, particularly for higher-performing CI users, whose baseline perceptual and cognitive skills are likely already robust. However, many CI users may be unable to attend to beneficial acoustic-phonetic cues in adverse listing conditions, such as in the presence of MTB, if top-down perceptual or cognitive skills are weak or if bottom-up auditory input is too impoverished to trigger such compensatory mechanisms, as in the CI-4 listener group. As a result, these CI users who perform worse under ideal conditions may suffer even greater declines in performance under challenging listening conditions compared to their better-performing counterparts.

ACKNOWLEDGMENTS

We thank Britt Bosma, Wilke Bosma, Roos van Doorn, and Anne Nijman for their assistance with this project. The study was supported in part by a VENI Grant (No. 275-89-035) from the Netherlands Organization for Scientific Research (NWO) and a VICI Grant (No. 918-17-603) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), and funds from the Heinsius Houbolt Foundation. The study is part of the research program of the Otorhinolaryngology Department of the University Medical Center Groningen: Healthy Aging and Communication.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**, 255–278.

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

Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (**2016b**). "Perception and psychoacoustics of speech in cochlear implant users," in *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine*, edited by A. T. Cacace, E. de Kleine, A. Holt, and P. van Dijk (Plural Publishing, Inc., San Diego, CA).

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top-down restoration of speech in cochlear-implant users," Hear. Res. **309**, 113–123.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2016**). "The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners," J. Assoc. Res. Otolaryngol. **17**, 475–491.

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P. H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., and Lazard, D. S. (**2013**). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients," Audiol. Neurotol. **18**, 36–47.

Calandruccio, L., Bradlow, A., and Dhar, S. (**2014**). "Speech-on-speech masking with variable access to the linguistic content of the masker speech for native and nonnative English speakers," J. Am. Acad. Audiol. **25**, 355–366.

Calandruccio, L., Van Engen, K., Dhar, S., and Bradlow, A. (**2010**). "The effectiveness of clear speech as a masker," J. Speech Lang. Hear. Res. **53**, 1458–1471.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

El Boghdady, N., Gaudrain, E., and Başkent, D. (**2019**). "Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users?," J. Acoust. Soc. Am. **145**, 417–439.

Ernestus, M., Hanique, I., and Verboom, E. (**2015**). "The effect of speech situation on the occurrence of reduced word pronunciation variants," J. Phon. **48**, 60–75.

Ernestus, M., and Warner, N. (**2011**). "An introduction to reduced pronunciation variants," J. Phon. **39**, 253–260.

Faulkner, K. F., Tamati, T. N., Gilbert, J. L., and Pisoni, D. B. (**2015**). "List equivalency of PRESTO for the evaluation of speech recognition," J. Am. Acad. Audiol. **26**, 582–594.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Gaudrain, E., and Başkent, D. (**2015**). "Factors limiting vocal-tract length discrimination in cochlear implant simulations," J. Acoust. Soc. Am. **137**(3), 1298–1308.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (**2007**). "Effect of spectral smearing on the perceptual segregation of vowel sequences," Hear. Res. **231**, 32–41.

Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (**2008**). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," J. Acoust. Soc. Am. **124**(5), 3076–3087.

Gilbert, J. L., Tamati, T. N., and Pisoni, D. B. (**2013**). "Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test," J. Am. Acad. Audiol. **24**, 26–36.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species – 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., and Brungart, D. (**2018**). "Clear speech adaptations in spontaneous speech produced by young and older adults," J. Acoust. Soc. Am. **144**, 1331–1346.

Helfer, K. S. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech," J. Speech, Lang. Hear. Res. **40**, 432–443.

Iverson, P., and Bradlow, A. R. (**2002**). "The recognition of clear speech by adult cochlear implant users," in *ICSA Workshop Temporal Integration in the Perception of Speech*, Aix-en Provence, France (April 8–10).

Janse, E., and Ernestus, M. (**2011**). "The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing," J. Phon. **39**, 330–343.

Janse, E., Nooteboom, S. G., and Quené, H. (**2007**). "Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition," Lang. Cogn. Process. **22**, 161–200.

Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., Skarzynski, P. H., Skarzynski, H., Schauwers, K., O'Leary, S., Mawman, D., Maat, B., Kleine-Punte, A., Huber, A. M., Green, K., Govaerts, P. J., Fraysse, B., Dowell, R., Diller, N., Burke, E., Beynon, A., Bergeron, F., Başkent, D., Artières, F., and Blamey, P. J. (**2012**). "Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time," PLoS One **7**, e48739.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F.-G. (**2004**). "Clear speech perception in acoustic and electric hearing," J. Acoust. Soc. Am. **116**, 2374–2383.

Luo, X., Fu, Q.-J., Wu, H.-P., and Hsu, C.-J. (**2009**). "Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users," Hear. Res. **256**(1), 75–84.

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (**2012**). "Speech recognition in adverse conditions: A review," Lang. Cogn. Process. **27**, 953–978.

Moberly, A. C., Lowenstein, J. H., and Nittrouer, S. (**2016**). "Word recognition variability with cochlear implants: 'perceptual attention' versus 'auditory sensitivity'," Ear Hear. **37**, 14–26.

Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., and Nittrouera, S. (**2014**). "Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing?," J. Speech, Lang. Hear. Res. **57**, 566–582.

Schoof, T., and Rosen, S. (**2014**). "The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners," Front. Aging Neurosci. **6**, 307.

Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol. **7**, 212–218.

Smiljanic, R., and Sladen, D. (**2013**). "Acoustic and semantic enhancements for children with cochlear implants," J. Speech Hear. Res. **56**, 1085–1096.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**(2), 1081–1091.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Tamati, T. N., Janse, E., and Başkent, D. (**2019**). "Perceptual discrimination of speaking style under cochlear implant simulation," Ear Hear. **40**, 63–76.

Van Engen, K. J., Chandrasekaran, B., and Smiljanic, R. (**2012**). "Effects of speech clarity on recognition memory for spoken sentences," PLoS One **7**, e43753.

Van Engen, K., Phelps, J., Smiljanic, R., and Chandrasekaran, B. (**2014**). "Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker," J. Speech Lang. Hear. Res. **57**, 1908–1918.

van Son, R. J. J. H., Binnenpoorte, D., Van Den Heuvel, H., and Pols, L. C. W. (**2001**). "The IFA corpus: A phonemically segmented Dutch 'open source' speech database," in *Proceedings of Eurospeech 2001*, Aalborg, Denmark.

van Son, R. J. J. H., Wesseling, W., Sanders, E., and Van Den Heuvel, H. (**2008**). "The IFADV corpus: A free dialog video corpus," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC* 2008, Marrakech, Morocco.