Perceptual Discrimination of Speaking Style Under Cochlear Implant Simulation

Terrin N. Tamati,[1,2] Esther Janse,[3] and Deniz Başkent[1,2]

Objectives: Real-life, adverse listening conditions involve a great deal of speech variability, including variability in speaking style. Depending on the speaking context, talkers may use a more casual, reduced speaking style or a more formal, careful speaking style. Attending to fine-grained acoustic-phonetic details characterizing different speaking styles facilitates the perception of the speaking style used by the talker. These acoustic-phonetic cues are poorly encoded in cochlear implants (CIs), potentially rendering the discrimination of speaking style difficult. As a first step to characterizing CI perception of real-life speech forms, the present study investigated the perception of different speaking styles in normal-hearing (NH) listeners with and without CI simulation.

Design: The discrimination of three speaking styles (conversational reduced speech, speech from retold stories, and carefully read speech) was assessed using a speaking style discrimination task in two experiments. NH listeners classified sentence-length utterances, produced in one of the three styles, as either formal (careful) or informal (conversational). Utterances were presented with unmodified speaking rates in experiment 1 (31 NH, young adult Dutch speakers) and with modified speaking rates set to the average rate across all utterances in experiment 2 (28 NH, young adult Dutch speakers). In both experiments, acoustic noise-vocoder simulations of CIs were used to produce 12-channel (CI-12) and 4-channel (CI-4) vocoder simulation conditions, in addition to a no-simulation condition without CI simulation.

Results: In both experiments 1 and 2, NH listeners were able to reliably discriminate the speaking styles without CI simulation. However, this ability was reduced under CI simulation. In experiment 1, participants showed poor discrimination of speaking styles under CI simulation. Listeners used speaking rate as a cue to make their judgements, even though it was not a reliable cue to speaking style in the study materials. In experiment 2, without differences in speaking rate among speaking styles, listeners showed better discrimination of speaking styles under CI simulation, using additional cues to complete the task.

Conclusions: The findings from the present study demonstrate that perceiving differences in three speaking styles under CI simulation is a difficult task because some important cues to speaking style are not fully available in these conditions. While some cues like speaking rate are available, this information alone may not always be a reliable indicator of a particular speaking style. Some other reliable speaking styles cues, such as degraded acoustic-phonetic information and variability in speaking rate within an utterance, may be available but less salient. However, as in experiment 2, listeners' perception of speaking styles may be modified if they are constrained or trained to use these additional cues, which were more reliable in the context of the present study. Taken together, these results suggest that dealing with speech variability in real-life listening conditions may be a challenge for CI users.

Key words: Cochlear implants, Speech perception, Speech variability.

(Ear & Hearing 2019;40;63–76)

INTRODUCTION

The adverse listening conditions commonly encountered in daily life can present significant challenges to successful speech communication. Real-life listening conditions may involve background noise or other masking speech, such as competing talkers, but also substantial variation intrinsic to the speech signal (Mattys et al. 2012). In real-life situations, listeners encounter talkers with diverse backgrounds in different contexts. As such, they hear multiple pronunciations of a word, which may differ across talkers and social groups, as well as environmental and social contexts.

Speech variability plays an important role in speech perception and spoken word recognition (Abercrombie 1967). Listeners simultaneously process both the linguistic information (e.g., words) and nonlinguistic information (e.g., talker characteristics) in the speech signal to be able to both understand the linguistic content of the utterance and use the nonlinguistic information to make judgements about the talker or context (Johnson & Mullennix 1997; Pisoni 1997). Listeners can use nonlinguistic information about talkers or contexts as a source of information to facilitate speech communication, for example, by attending to acoustic-phonetic cues characterizing a type of speech variability and matching patterns to stored representations of that variability. This process allows listeners to make nonlinguistic judgments about speech, including the talker's identity (Van Lancker et al. 1985a, b), gender (Lass et al. 1976), age (Ptacek & Sander 1966), and region of origin and background (Labov 1972). Further, the linguistic and nonlinguistic information in the speech signal interact to influence speech recognition, as listeners are able to learn talker-, group-, or context-specific acoustic-phonetic patterns to adopt the most successful processing strategy and facilitate speech recognition (Nygaard et al. 1994; Nygaard & Pisoni 1998; Brouwer et al. 2012).

Speech variability may present a significant challenge for hearing-impaired users of cochlear implants (CIs), the auditory prosthetic devices for deaf people. Although CIs have been very successful as a medical treatment for profound deafness, the speech signal transmitted by a CI is less detailed in spectrotemporal cues compared with what is typically available to normal-hearing (NH) listeners (for a review, see Başkent et al. 2016). The degraded speech signal does not convey the fine phonetic details in speech, including both the acoustic-phonetic properties representing linguistic contrasts in their language and the

[1]Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [2]Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands; and [3]Centre for Language Studies, Radboud University Nijmegen, Nijmegen, The Netherlands.

speech variability that characterizes different talkers, accents, or contexts. The acoustic-phonetic cues that are especially important for the perception of different sources of speech variability are degraded and underspecified compared with the more robust information available to NH listeners. For example, some recent studies have found that CI users show poor perception of cues that are important for talker and gender discrimination in NH listeners, including fundamental frequency (F0) and vocal-tract length (Fu et al. 2004; Green et al. 2004; Laneau et al. 2004; Moore & Carlyon 2005; Fuller et al. 2014; Gaudrain & Başkent 2018).

As a result of the limited speech variability information, CI users may be less able to use speech variability cues to facilitate both understanding the linguistic content of the utterance and making nonlinguistic judgements about the talker or context. The speech recognition skills of CI users are generally quite high for ideal speech materials, that is, carefully articulated speech produced by a single talker. Yet, for many CI users, speech recognition is poor for high-variability materials (Gilbert et al. 2013; Faulkner et al. 2015) or real-life challenging forms of speech (e.g., conversational speech, Liu 2014; fast speech, Ji et al. 2013; accented speech, Ji et al. 2014). Similarly, some previous studies have shown that CI users have difficulty discriminating different talkers (Cleary & Pisoni 2002; Cleary et al. 2005) and genders (Massida et al. 2011). Additionally, CI users are less sensitive to differences between foreign-accented and native speech (Tamati & Pisoni 2015) and differences among regional accents (Clopper & Pisoni 2004a; Tamati et al. 2014). In these studies, CI users were able to detect some differences in talkers' voices and accents. However, they consistently perceived smaller differences between the foreign-accented and native speech and among different regional accents, compared with NH listeners. Thus, while some acoustic-phonetic information characterizing different sources of speech variability may be available, the amount of information CI users receive is likely degraded and underspecified compared with the more robust information NH listeners can perceive and encode. Because of this, CI users may not be able to rely on the same cues as NH listeners in their perception of speech variability. Further, they may be overall much less able to make use of available speech variability information to support speech recognition.

Little is known about how much or what type of speech information is available to CI users with the existing processing strategies, and how they use this available information to be able to perceive and understand variations of natural speech. The present study explores these issues by investigating the perception of speaking style by NH listeners under acoustic simulations of CI speech processing. Speaking style is a common source of real-life speech variability. Depending on the speaking environment or goal of a speaking task, talkers use different speaking styles, such as very carefully produced, clearly articulated speech or very casually produced, reduced speech (Ernestus et al. 2015). For example, a common form of reduction in English and Dutch is /t/-deletion after /s/ (particularly before /b/, Mitterer & Ernestus 2006; Janse et al. 2007; Ernestus et al. 2015). In a careful pronunciation of the word *kast* (cabinet) in *kast brengen* (bring a cabinet), the final /t/ of *kast* may be produced with a release burst (e.g., *kast brengen*), while in a reduced pronunciation, the final stop may be unreleased (e.g., *kas_*). Speaking style variability presents a good tool for exploring CI users' implantation outcome because speaking styles differ in a range of both spectral and temporal cues. Careful speech is characterized by hyperarticulated sound segments and syllables, increased loudness, and a relatively slow speaking rate. In conversational reduced speech, speech sounds are often shorter, weaker, or even absent, and speaking rate is often faster and more variable compared with more fully articulated careful speech. While spectral information is poorly encoded by the CI device (Friesen et al. 2001), temporal information is more robustly encoded (Moore & Glasberg 1988; Shannon 1992; Fu et al. 2004; Nie et al. 2006; Gaudrain & Başkent 2015). As such, important cues characterizing a particular speaking style conveyed by fine spectral detail may not be fully available to CI users, which may force listeners to rely more on cues conveyed by temporal information (Fuller et al. 2014; Tamati et al. 2014).

In the present study, the perception of three speaking styles, specifically carefully read speech, speech from retold stories, and speech produced in the context of a conversation, was assessed in a speaking style discrimination task with speech with natural speaking rates (experiment 1) and with modified speaking rates (experiment 2). Listeners were presented with a single utterance and were asked to determine the style of speech that the talker used to produce that utterance. In this task, speaking style perception relies not only on the discrimination or detection of acoustic-phonetic information related to different speaking styles in the speech signal but also to long-term knowledge of speaking styles. Listeners must be able to perceive speaking style-specific cues in a target utterance, and use previously acquired knowledge of the acoustic-phonetic characteristics of different speaking styles to associate the target utterance to a particular speaking style. Because speaking style perception relies at least partially on prior knowledge and experience, speech from three different contexts was selected to obtain a range of speaking styles that may reflect variability encountered in real-life communicative contexts. Further, previous research on reduction across many contexts suggests that speaking style differences are most evident between scripted speech (i.e., read speech) and nonscripted speech (i.e., speech formulated on the spot), and that there are additional differences within nonscripted speech depending on the formality of the speaking context (Ernestus et al. 2015). Therefore, in the present study, read speech represents more carefully articulated speech, while speech produced in a conversation represents more casually articulated reduced speech and retold stories represent speech that is neither very carefully nor casually articulated. All materials were selected to contain similar, but not identical, linguistic content (i.e., discussion of vacation) across speaking styles. With these materials, listeners must rely on many different speaking style cues, rather than specific cues related to a particular word or phrase in speech with controlled linguistic content (Clopper & Pisoni 2004b) or words or phrases common to informal (e.g., weather) or formal (e.g., formal texts) settings with unrestricted linguistic content. Thus, the speech materials used in the present study allowed us to assess listeners' perception of a broad range of speaking style cues and to obtain a general measure of their ability to perceive differences in speaking styles.

In the discrimination task in experiment 1, utterances were presented without CI simulation and with 12- and 4-channel CI simulation, approximating the range of excellent and poor CI hearing, respectively (Friesen et al. 2001), and facilitating the assessment of the influence of spectral resolution on speaking style perception. Since many acoustic-phonetic cues indexing

speaking style are carried in the spectral properties of the vowels and consonants, performance was expected to decline with applying both CI simulation and decreasing spectral resolution. However, if listeners rely heavily on temporal cues, such as speaking rate, to discriminate speaking style in the absence of rich spectral detail, then performance would not drastically differ between CI-simulation conditions. Experiment 2 investigated the role of speaking rate in speaking style perception. Previous research suggests that CI users are sensitive to differences in speaking rate. In particular, speech produced with faster speaking rates has been shown to be more challenging to recognize for CI users and NH listeners tested under CI simulation (Ji et al. 2014; Jaekel et al. 2017). These findings suggest that speaking rate may be a potentially useful cue in the speaking style discrimination task. In experiment 2, we used a manipulation of speaking rate, in addition to the CI simulations, to minimize differences in speaking rate among the three speaking styles. If listeners rely heavily on speaking rate information, especially in the CI-simulation conditions, the discrimination of speaking styles would be poor. Thus, together, the manipulations in experiments 1 and 2 allowed us to explore the contributions of spectral and temporal cues in speaking style perception with and without CI simulation.

EXPERIMENT 1

Methods

Participants • Thirty-one (27 female, 4 male) NH young adults participated in experiment 1. Participants were native speakers of Dutch between the ages of 18.7 and 24.6 years (M = 21.8 years). All passed a pure-tone hearing screening test at 20 dB HL from 250 to 8000 Hz for both ears, and none reported a history of hearing or speech disorders at the time of testing. Participants received 8 euros for 1 hr of testing. The experiment was approved by the ethics committee of the University Medical Center Groningen (METc 2012.455).

Materials

Three female (20, 28, 60 years old) and 3 male (40, 56, 66 years old) talkers were selected from the Instituut voor Fonetische Wetenschappen Amsterdam (IFA) corpus of the Institute of Phonetic Sciences Amsterdam (IFA; Van Son et al. 2001). All talkers were native speakers of Dutch, with varying regions of origin. Two talkers (1 female/1 male) were born and attended primary and secondary school in the West of the Netherlands (Zeeland, Noord Holland), 2 talkers (1 female/1 male) in the East of the Netherlands (Overijssel, Gelderland), and 2 talkers (1 female/1 male) in more than 1 region (Gelderland-Noord-Brabant, Friesland-Gelderland). These talkers were selected because of the quality of the recordings and for the number of sentence-length utterances that met the criteria described below.

For the speaking style discrimination task, materials consisted of 162 unique sentence-length utterances, 27 for each of the 6 talkers. For each talker, hence, 9 utterances were from the context of a conversation (casual conversation), 9 utterances from a retelling of a story (retold story), and 9 utterances from a read list (careful read). The target utterances were selected to obtain similar semantic and syntactic content across the three speaking styles. All utterances concerned the details of a vacation, minimizing the general content to be potentially used as an aid in the discrimination task. Similarly, the number of words in the utterances varied within each speaking style category, but did not differ substantially across speaking styles (mean number of words per utterance in the discrimination task: casual conversation = 8.8 [SD = 2.6], retold story = 8.8 [SD = 2.1], careful read = 8.4 [SD = 2.4]). Detailed analyses of the characteristics of the stimulus materials are provided below.

Sentences were presented in one of three simulation conditions: (1) unprocessed (no simulation), (2) a 12-channel CI simulation (CI-12), or (3) a 4-channel CI simulation (CI-4). For the simulation conditions, sentences were processed through a 12- or 4-channel noise-band vocoder implemented in Matlab. This was achieved by filtering the original signal into 12 or 4 bands between 150 and 7000 Hz, using 12th order, zero-phase Butterworth filters. The bands were partitioned based on Greenwood's frequency-to-place mapping function, simulating evenly spaced regions of the cochlea (Greenwood 1990). The same cutoff frequencies were used for both the analysis and synthesis filters. From each frequency band, the temporal envelope was extracted by half-wave rectification and low-pass filtering at 300 Hz, using a zero-phase 4th order Butterworth filter. Noise-band carriers were generated independently for each channel by filtering white noise into spectral bands using the same 12th order Butterworth band-pass filters. The final 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.

To distribute the sentences across the three CI-simulation conditions across participants, each sentence was randomly assigned to one of three lists A, B, and C. Each list contained a total of 54 sentences (9 per talker), with 18 for each of the three speaking styles (3 per talker). The presentation order of the lists was balanced across participants, with 6 different presentation orders (order 1, n = 5: A, B, C; order 2, n = 4: A, C, B; order 3, n = 6: B, A, C; order 4, n = 6: B, C, A; order 5, n = 5: C, A, B; order 6, n = 5: C, B, A).

To ensure that recognition accuracy was not at or near floor in the difficult 4-channel CI-simulation condition, a speaking style sentence recognition task was used to assess the intelligibility of the three speaking styles and three simulation conditions. A set of 54 unique sentence-length utterances (9 for each of the same 6 talkers, with 3 sentences from each speaking style per talker) was selected for a speaking style intelligibility task. To distribute the sentences across the three simulation conditions across participants, each sentence was again randomly assigned to one of three lists A, B, and C. Each list contained a total of 18 sentences, 6 for each of the 3 speaking styles (1 per talker). The presentation order for the sentence recognition task was matched to the discrimination task, such that each participant assigned to order 1 for the discrimination task was also assigned to order 1 for the sentence recognition task, and so on (except for one participant who completed order 1 for the discrimination task and order 2 for the sentence recognition task).

Analyses of Stimulus Materials • To assess the extent to which the selected materials used in the speaking style discrimination task represent three distinct speaking styles, we conducted a series of analyses on the stimulus materials. Several acoustic-phonetic characteristics were selected, based on previous accounts of the differences between clear and conversational speech in Dutch (Ernestus 2000; Schuppler et al. 2011). All measurements were collected from the actual stimulus materials used in the speaking style discrimination task.

TABLE 1. Stimulus Properties of CC, RS, and CR Materials: Summary of Stimulus Properties for the 3 Female and 3 Male Talkers

	Female									Male								
	F20			F28			F60			M40			M56			M66		
Stimulus Properties	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR
Total number of disfluencies (uh, um)	1	3	0	0	1	0	1	2	0	2	3	0	3	2	0	3	0	0
Total number of speech errors	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total number of informal words (ja, maar, nou, nee)	1	2	0	1	2	0	1	1	0	3	0	1	3	0	1	2	2	2

CC, casual conversation; RS, retold story; CR, careful read.

Comparisons were carried out between the careful read, retold story, and casual conversation speech for each of the 6 talkers and across all 6 talkers.

Table 1 displays the word types used in each of the three speaking styles for each talker, and Table 2 displays the word types used in the three speaking styles collapsed across talkers. Disfluencies (e.g., uh, um), speech errors, and informal words (e.g., ja [yes], maar [but], nou [well], nee [no]) are more common in conversational speech (Schuppler et al. 2011). Overall, in our materials, there were few disfluencies (n = 21) and speech errors (n = 1) across all talkers and speaking styles. No disfluencies or speech errors were present in the careful read speech, and the rest were equally distributed in the retold story (n = 11) and casual conversation (n = 11) speech. A total of 22 informal words, common in conversational speech, were present in the stimulus materials across all talkers and speaking styles. Casual conversation speech contained 11 informal words, retold story speech contained 7 informal words, and careful read speech contained only 4 informal words. The analysis on word types, although low in number, suggests that casual conversation speech contained more word types characteristic of conversational speech, and careful read speech contained few word types characteristic of conversational speech. Retold story also contained some word types characteristic of conversational speech (disfluencies, speech errors), but fewer overall.

A series of acoustic analyses were also carried out on the stimulus materials, for each talker (Table 3) and collapsed across talkers (Table 4). The number of pauses present in the stimulus items was calculated. A pause was defined as a period of silence longer than 200 msec. More pauses are associated with a careful speaking style (Bradlow et al. 2003). Overall, there were 17 pauses. The most pauses were produced in the retold story (n = 12) and careful read (n = 3) speech, compared with the casual conversation (n = 2) speech. The number of pauses in the retold story speech, but not careful read speech, was more consistent with a more careful speaking style. However, because there were very few pauses overall, with individual talkers producing only one to five pauses, the number of pauses was not a defining characteristic of the speaking styles in these materials.

TABLE 2. Stimulus Properties of CC, RS, and CR Materials: Summary of Stimulus Properties Collapsed Across Talkers

Stimulus Properties	CC	RS	CR
Total number of disfluencies (uh, um)	10	11	0
Total number of speech errors	1	0	0
Total number of informal words (ja, maar, nou, nee)	11	7	4

CC, casual conversation; RS, retold story; CR, careful read.

The average speaking rate (including pauses) for each condition was measured in realized syllables per second using a Praat script (de Jong & Wempe 2009). A faster speaking rate is generally a feature of a conversational speaking style, while a slower speaking rate is a feature of a careful speaking style (Bradlow et al. 2003). The average speaking rate across all materials was 4.20 syllables/s. Casual conversation speech was produced slightly faster (M = 4.36 syllables/s) than both retold story (M = 4.05 syllables/s) and careful read (M = 4.18 syllables/s). Thus, in terms of speaking rate, casual conversation speech was consistent with a more conversational speaking style, and retold story and careful read speech were more consistent with a more careful speaking style. However, this pattern varied substantially across talkers; 3 talkers demonstrated faster speaking rates in casual conversation speech compared with careful read speech (F60, M56, M66), but 3 talkers demonstrated faster or similar speaking rates for retold story speech compared with casual conversation speech (F20, F28, M40). Additionally, the retold story speech was more consistent with careful read speech for only 2 talkers (M56, M66), with casual conversation speech for only 1 talker (F20), with neither speaking styles for 3 talkers (F28, F60, M40).

Increased average pitch and range has been found to be characteristic of a more careful speaking style in English (Picheny et al. 1986; Krause 2001; Bradlow et al. 2003). The average F0 across all talkers was slightly higher in the careful read speech (M = 159.1 Hz) than the retold story (M = 156.9 Hz) and the casual conversation (M = 148.1 Hz) speech. Similarly, the F0 range (measured in SD) was greater in the careful read (SD = 30.2 Hz) than the retold story (SD = 26.3 Hz) and the casual conversation (SD = 25.0 Hz) speech. Thus, in terms of F0 average and range, the careful read speech is more characteristic of a careful speaking style, and the casual conversation speech is more characteristic of a conversational speaking style, while the retold story speech is between the other two.

Finally, we examined some specific phenomena that have generally been found to be useful in describing different speaking styles in Dutch (Ernestus 2000; Schuppler et al. 2011). We broadly examined the deletion of [t] in word-final position after a consonant, the deletion of schwa in unstressed syllables, the

TABLE 3. Acoustic Analysis of CC, RS, and CR Materials: Summary of Acoustic Analysis for the 3 Female and 3 Male Talkers

	Female									Male								
	F20			F28			F60			M40			M56			M66		
Acoustic Measurements and Observations	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR	CC	RS	CR
Total number of pauses	0	3	0	1	0	0	0	4	1	0	3	0	0	0	2	1	2	0
Average speaking rate (number syllables/s)	4.11	3.98	4.58	4.50	4.87	4.52	4.53	4.28	3.98	3.75	3.27	3.82	4.43	3.92	4.22	4.81	3.99	3.98
F0 mean (Hz)	195.1	204.9	186.3	186.4	204.3	206.1	169.7	182.9	190.1	96.8	104.8	107.4	106.7	103.7	111.9	133.6	141.1	152.8
F0 range in SD (Hz)	28.5	33.1	31.9	29.0	33.9	36.6	31.1	29.9	38.5	11.9	17.3	15.1	27.63	18.2	24.5	21.6	25.4	34.3
Average rate (%) of word-final [t]-realization	70.0	40.0	83.3	100.0	60.0	83.3	50.0	62.5	75.0	76.9	71.4	100.0	71.4	71.4	88.9	33.3	71.4	100.0
Average rate (%) of schwa realization in unstressed syllables	64.5	74.1	88.5	79.3	93.6	100.0	88.9	92.3	94.6	70.0	90.9	90.3	70.4	83.3	88.9	89.5	100.0	100.0
Average rate (%) of word-final [n]-realization	10.0	50.0	62.5	42.9	100.0	84.6	40.0	28.6	40.0	14.3	28.6	46.2	12.5	53.9	40.0	66.7	83.3	83.3
Average rate (%) of postvocalic [r]-realization	60.0	71.4	100.0	50.0	66.7	80.0	100.0	100.0	85.7	20.0	100.0	80.0	40.0	100.0	66.7	66.7	100.0	100.0

CC, casual conversation; RS, retold story; CR, careful read.

TABLE 4. Acoustic Analysis of CC, RS, and CR Materials: Summary of Acoustic Analysis Collapsed Across Talkers

Acoustic Measurements and Observations	CC	RS	CR
Total number of pauses	2	12	3
Average speaking rate (number syllables/s)	4.36	4.05	4.18
F0 mean (Hz)	148.1	156.9	159.1
F0 range in SD (Hz)	25.0	26.3	30.2
Average rate (%) of word-final [t]-realization	69.8	61.4	87.2
Average rate (%) of schwa realization in unstressed syllables	26.7	53.5	38.0
Average rate (%) of word-final [n]-realization	77.9	89.4	93.9
Average rate (%) of postvocalic [r]-realization	52.5	87.1	84.4

CC, casual conversation; RS, retold story; CR, careful read.

deletion of [n] in word-final position, and the deletion of [r] in a postvocalic position. These sound segments are commonly deleted in a more conversational speaking style compared with a more careful speaking style. The average rates of the realization of these sound segments were calculated by counting the total number of these segments audibly present compared with the total number of possible occurrences. In the current speech materials, the talkers tended to produce the [t] in word-final position after a consonant more often in careful read speech (87.2%) than both retold story (61.4%) and casual conversation (69.8%) speech. The talkers produced the schwa in unstressed syllables more often in careful read speech (93.9%) than in retold story (89.4%) and casual conversation (77.9%) speech. Similarly, talkers produced the [n] in word-final position more often in careful read speech (62.0%) than in retold story (52.5%) and casual conversation (26.7%) speech. They produced the [r] in postvocalic position more often in careful read (84.4%) and retold story (87.1%) speech than in casual conversation (52.2%) speech. Taken together, overall, the selected sound segments were more often fully realized in careful read speech than in casual conversation speech. Further, the tendency to fully realize these sound segments in careful read speech was also fairly consistent across talkers. Therefore, in terms of these deletion phenomena, the careful read speech is more characteristic of a careful speaking style, and the casual conversation speech is more characteristic of a conversational speaking style, with retold story speech displaying rates between the other two.

The analyses of the stimulus materials used in the discrimination task have shown that overall the careful read speech produced by the 6 talkers displays more properties consistent with a careful speaking style and the casual conversation speech displays more properties consistent with a conversational speaking style. The retold story speech displayed properties of either careful read or casual conversation speech for some measures, and for many measures presented an in-between case. Thus, the speech from these three categories seemed to present a range of speaking styles (and speaking style cues) and is largely consistent with previous characterizations of speaking style differences among scripted speech and different variations of nonscripted speech (Ernestus et al. 2015). Therefore, it was

expected that there were many spectral and temporal cues that could be potentially used to discriminate the speaking styles in the discrimination task.

Procedure

Participants were tested individually, seated in a sound-attenuated booth in front of a computer monitor. The computer-based tasks were run on MacOS X, using experimental programs controlled by PsyScope X B77 scripts. Stimulus materials were presented binaurally through HD600 headphones (Sennheiser GmbH & Co., Wedemark, Germany), via an AudioFire4 soundcard (Echo Digital Audio Corp, Santa Barbara, CA) that was connected to a DA10 D/A converter (Lavry Engineering, Poulsbo, WA). Output levels of the target sentences were calibrated to be approximately 65 dB sound pressure level.

Speaking Styles Discrimination Task • On each trial, participants were presented with a single unique utterance, with no repetition. The participants were asked to respond if the utterance was produced in a formal (careful) manner or an informal (casual or conversational) manner by pressing a button on the keyboard corresponding to formal/careful ("1") or informal/casual ("0"). Participants were told to wait until the end of the utterance to respond, and then give their response as quickly as possible without compromising accuracy. The participants had 3 sec to respond before the next trial began.

Before testing, participants were given written and oral descriptions and examples of carefully and casually articulated speech by the experimenter. To familiarize the participants with the task, one practice list of 6 unprocessed items was used. The practice items consisted of 3 casual conversation and 3 careful read utterances produced by 2 additional talkers (1 female/1 male) from the IFA corpus (Van Son et al. 2001) who were not selected for the test items. All participants were presented with unprocessed utterances from the first experimental list (no simulation), 12-channel CI-simulated utterances from the second experimental list (CI-12), and 4-channel CI-simulated utterances from the third experimental list (CI-4). Within each experimental list, utterances were randomly presented. A break was given between each block.

Trial responses were coded for speaking style (careful or casual) and simulation condition. Trials for which no response was given within 3 sec were excluded from the analysis (approximately 2%, n = 79/5022).

Speaking Styles Intelligibility Task • Participants were presented with a single unique utterance, with no repetition, on each trial. The participants were asked to type the words that they heard using the keyboard. Partial answers and guessing were encouraged. To familiarize the participants with the task, one practice list of six unprocessed sentences preceded testing. Again, the practice items consisted of three casual conversation and three careful read utterances produced by 2 additional talkers (1 female/1 male) from the IFA corpus (Van Son et al. 2001), who were the same talkers from the practice trials in the discrimination task. All participants were presented with unprocessed utterances from the first experimental list (no simulation), 12-channel CI-simulated utterances from the second experimental list (CI-12), and 4-channel CI-simulated utterances from the third experimental list (CI-4). Within each experimental list, utterances were randomly presented. The experiment was self-paced, without time limits, and breaks were encouraged between lists.

Scoring was completed off-line for number of words correct. Exact word order was not required, but plural or possessive morphological markers were required to match the word. Minor spelling errors were also accepted as long as the error did not result in an entirely different word. The total number of words correctly recognized was collected and analyzed for each speaking style and simulation condition.

Results

Speaking Style Discrimination Task • To assess whether listeners were able to make explicit judgments about speaking style, the percent careful and casual responses were examined by speaking style. For each participant, the percent of careful responses was calculated by dividing the number of trials for which he/she gave a careful response by the total number of trials in a particular condition. In the no-simulation condition, participants gave more careful ratings for the careful read speech and more casual ratings for the casual conversation speech. Retold story speech was rated nearly equally as careful and casual. Figure 1 shows the median percent careful responses for careful read, retold story, and casual conversation utterances.

A repeated measures analysis of variance (ANOVA) with speaking style (careful read, retold story, or casual conversation trials) and simulation condition (no simulation, CI-12, CI-4) as within-subject factors was carried out on the listeners' percent careful responses. The analysis revealed a significant main effect of speaking style [$F(2, 60) = 30.95$; $p < 0.001$] and a significant interaction of simulation condition and speaking style [$F(4, 120) = 3.94$; $p = .005$]. The main effect of simulation condition was not significant. To examine the effect of speaking style in further detail, a series of post hoc paired comparison t tests were carried out on the responses for each speaking style; no corrections for multiple comparisons were applied. Careful read speech received significantly more careful judgements than retold story [$t(30) = 2.86$; $p = 0.008$] and casual conversation speech [$t(30) = 6.92$; $p < 0.001$], and retold story speech was rated as more careful than casual conversation speech [$t(30) = 5.23$; $p < 0.001$]. An additional set of t tests on speaking styles responses for each simulation condition was carried out to explore the interaction of speaking styles and simulation condition. For the no-simulation condition, careful read speech received significantly more careful judgements than retold story [$t(30) = 3.73$; $p = 0.001$] and casual conversation speech [$t(30) = 8.24$; $p < 0.001$], and retold story speech was rated as more careful than casual conversation speech [$t(30) = 4.55$; $p < 0.001$]. For the CI-12 condition, none of the comparisons reached significance. For the CI-4 condition, careful read speech received significantly more careful judgements than casual conversation speech [$t(30) = 2.98$; $p = 0.006$] and retold story speech was rated as more careful than casual conversation speech [$t(30) = 2.30$; $p = 0.029$], but no other comparison reached significance.

Speaking Style Intelligibility Task • To assess the impact of the CI simulations on the intelligibility of the three speaking styles, word recognition responses were examined for the three speaking styles across the three simulation conditions in the speaking style intelligibility task. Word recognition accuracy in the no-simulation condition was near ceiling with a mean accuracy of 94.8% (SD = 1.9), while performance declined in the CI-12 condition with a mean accuracy of 88.6% (SD = 3.9). Word recognition accuracy further declined in the CI-4 condition but was still at 61.5% (SD = 7.5).

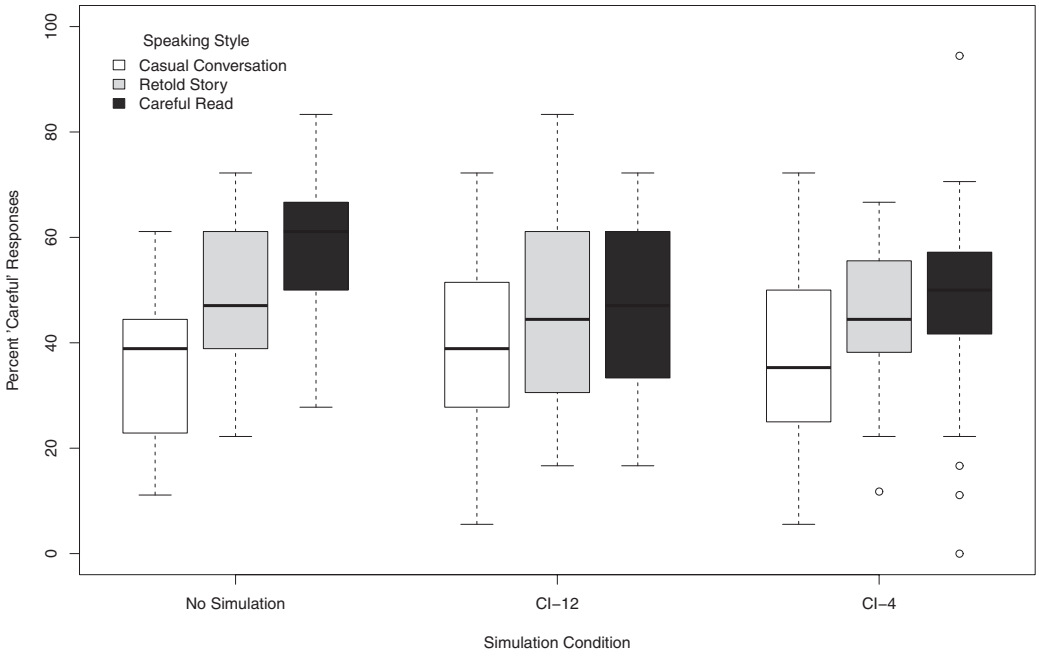

Figure 1. Box plot demonstrating the percent careful ratings from all listeners for careful read, retold story, and casual conversation utterances under all three simulation conditions (no simulation, CI-12, CI-4) in experiment 1. The boxes extend from the lower to the upper quartile (the interquartile range [IQ]), and the midline indicates the median. 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.

A repeated measures ANOVA with speaking style (careful read, retold story, or casual conversation trials) and simulation condition (no simulation, CI-12, CI-4) as within-subject factors was carried out on word recognition accuracy. Before analysis, the proportional word recognition accuracy scores were converted to rationalized arcsine units (Studebaker 1985) to account for differences in normality of the distributions. The analysis revealed a significant main effect of speaking style [$F(2, 60) = 140.00$; $p < 0.001$], a significant main effect of simulation condition [$F(2, 60) = 382.13$; $p < 0.001$], and a significant interaction of simulation condition and speaking style [$F(4, 120) = 4.92$; $p = 0.001$]. Word recognition accuracy for careful read, retold story, and casual conversation utterances under no simulation, CI-12, and CI-4 simulation conditions is provided in Figure 2.

A series of post hoc paired comparison t tests on simulation condition confirmed that accuracy in the no-simulation condition was greater than in the CI-12 condition [$t(30) = 9.41$; $p < 0.001$] and the CI-4 condition [$t(30) = 37.18$; $p < 0.001$]. Accuracy in the CI-12 condition was also greater than in the CI-4 condition [$t(30) = 16.96$; $p < 0.001$]. An additional set of paired comparison t tests on speaking style further revealed that overall careful read speech was more intelligible than retold story speech [$t(30) = 9.90$; $p < 0.001$] and casual conversation speech [$t(30) = 8.28$; $p < 0.001$], and casual conversation speech was more intelligible than retold story speech [$t(30) = 3.76$; $p = 0.001$].

The significant interaction of simulation condition and speaking style suggests that the effect of speaking styles on intelligibility was greater in some simulation conditions. For the no-simulation condition, careful read speech was more intelligible than retold story speech [$t(30) = 13.89$; $p < 0.001$] and casual conversation speech [$t(30) = 12.77$; $p < 0.001$], but casual conversation speech was not more intelligible than retold story speech. For the CI-12 condition, careful read speech was again more intelligible than retold story speech [$t(30) = 9.84$; $p < 0.001$] and casual conversation speech [$t(30) = 5.88$; $p < 0.001$], and casual conversation speech was more intelligible than retold story speech [$t(30) = 6.24$; $p < 0.001$]. For the CI-4 condition, careful read speech was more intelligible than retold story speech [$t(30) = 5.16$; $p < 0.001$] and casual conversation speech [$t(30) = 3.9$; $p < 0.001$], and casual conversation speech was more intelligible than retold story speech [$t(30) = 2.08$; $p = 0.046$].

Discussion

Experiment 1 investigated NH listeners' perception of speaking style with or without CI simulation. In the no-simulation condition, participants gave significantly more careful ratings for the careful read speech and more casual ratings for the casual conversation speech, with retold story speech rated equally as careful and casual. Thus, listeners were able to use meaningful and reliable acoustic-phonetic information related to speaking style in the speech signal to discriminate and categorize speaking style.

Listeners' perception of speaking style was worse under CI simulation than in the no-simulation condition. In the CI-12 condition, listeners did not distinguish any of the three speaking styles. In the CI-4 condition, careful read speech was rated as more careful than casual conversation speech, and retold story speech was rated as more careful than casual conversation speech. Thus, listeners may have been able to use some style-specific acoustic-phonetic differences to discriminate the three speaking styles but these differences were more limited than in the unprocessed condition. In addition, performance was not at or near floor in any condition in the speaking style intelligibility task. Thus, substantial linguistic information was still available

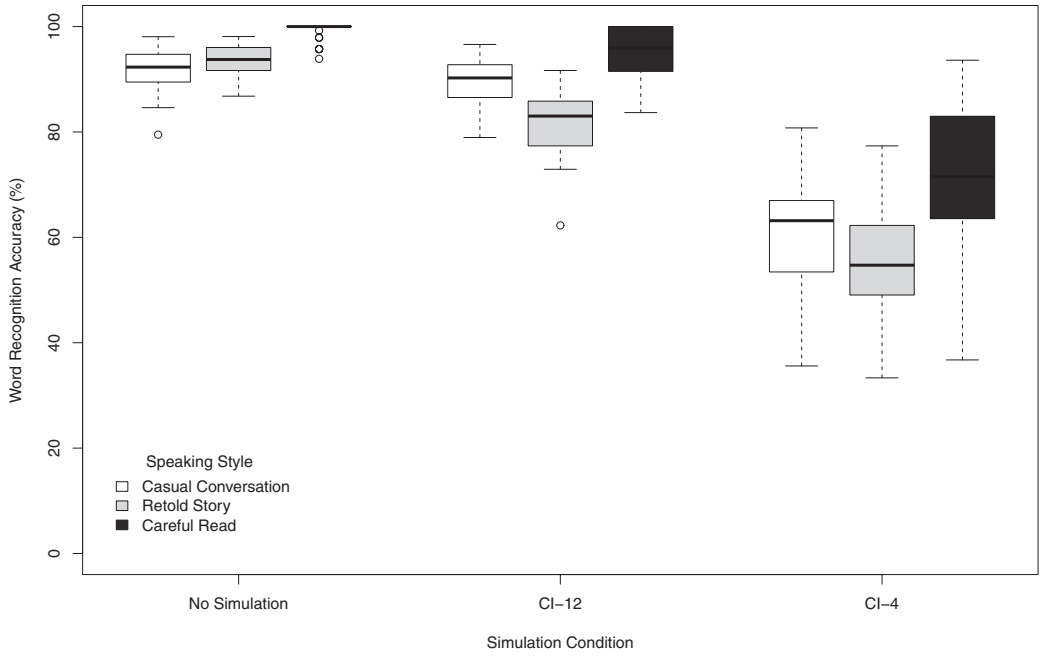

Figure 2. Word recognition accuracy in experiment 1 by speaking style (careful read, retold story, and casual conversation) and simulation condition (no simulation, CI-12, and CI-4). See Figure 1 for a description of the box plot design.

to the listener even in the most degraded conditions, and the poor discrimination of speaking style was more likely related to difficulties detecting detailed pronunciation differences rather than low intelligibility. This supports previous studies demonstrating that discrimination of different sources of speech variability suffers when reliable cues are limited by additional sources of degradation, like CI simulation (e.g., talker voice cues, Gaudrain & Başkent 2015) or noise (e.g., regional dialects, Clopper & Bradlow 2008).

CI simulation had an effect on the discrimination of the three speaking styles, but discrimination performance was poor in both CI-simulation conditions. We had expected that performance would also decline from the 12- to the 4-channel CI-simulation condition because many acoustic-phonetic cues indexing speaking style are carried in the spectral properties of the vowels and consonants. While listeners were unable to perceive differences between the speaking styles in the CI-12 condition, they perceived some minor differences among the speaking styles in the CI-4 condition. The speaking style cues that would have been available in all conditions were limited. Potential cues include word-level differences (Tables 1 and 2) in the number of disfluencies, speech errors, and possibly informal words, given that speech was largely intelligible in all simulation conditions, and differences in the number of pauses (Tables 3 and 4). As such, the few speaking style cues available in either CI-simulation condition may not have been sufficient to yield good discrimination of the speaking styles.

Based on the acoustic analyses, pitch (F0) differences and sound segment deletion differences (Tables 3 and 4) were reliable cues to potentially use to discriminate the speaking styles. However, while these cues would have been available in the unprocessed condition, they would have been greatly reduced in the CI-simulation conditions. Similar to the sound segment deletion differences, although not directly measured in the analyses of the stimulus materials, subtle pronunciation differences relating to sound segment reduction (e.g., vowel reduction to schwa and overall decreased vowel dispersion) would also be limited in the CI-simulation conditions. Correlational analyses between discrimination responses confirmed that F0 mean and range and the overall rate of sound segment deletion (calculated as the percent of the four target sounds in Tables 3 and 4 that were deleted in each utterance) was not significantly correlated with how often an item was categorized as careful in the CI-simulation conditions. F0 mean and range was significantly correlated with the percent careful ratings in the no-simulation condition (mean: $r = 0.26$, $p = 0.003$; range: $r = 0.23$, $p = 0.003$) but the rate of sound segment deletion was not. The significant relation between F0 mean and percent careful ratings likely reflects a tendency to rate the female talkers' speech as more careful in the no-simulation condition but not in the CI-simulation condition where this cue may not have been as salient. The relation between F0 range and the percent careful ratings reflects the tendency to categorize utterances with a greater range of F0 as careful, at least in the no-simulation condition. Sound segment deletion differences, and potentially, by extension, more subtle differences in sound segment reduction did not seem to contribute to the perception of speaking style in any condition.

Speaking rate information would also be available in the CI-simulation conditions, in addition to the word and pause differences. Correlational analyses between discrimination responses showed that speaking rate was significantly correlated with how often an item was categorized as careful without CI simulation ($r = -0.16$; $p = 0.043$), in the CI-12 condition ($r = -0.19$; $p = 0.016$), and in the CI-4 condition ($r = -0.21$; $p = 0.009$), with faster speaking rate being associated with fewer careful responses. Because they could not attend to more reliable speaking style information, listeners may have used speaking rate to make their judgements. However, in the current set of stimulus materials, although casual conversation utterances

were slightly faster than careful read utterances, the retold story was somewhat slower than the other categories. See Table 4 for the average speaking rate for each speaking style collapsed across talkers. Because speaking rate alone was not sufficient for accurate discrimination of the speaking styles with the current set of materials, this may have led to the listeners to make erroneous judgements of speaking style, resulting in the overall very poor discrimination scores in the discrimination task.

Experiment 2 was carried out to further explore the perception of speaking style with and without CI simulation. In experiment 1, speaking rate was used as a cue for making judgements about the speaking style of an utterance. However, speaking rate was not a reliable cue to speaking style in the current set of materials. By manipulating the speaking rate of the utterances, we examined the extent to which this cue influences speaking style perception with and without CI simulation and whether listeners are able to rely on other, potentially more reliable cues when variability in speaking rate is minimized. Utterances were temporally modified to have the same speaking rate, set as the average speaking rate (in syllables per second) of all utterances across all three speaking style categories. Based on these manipulations, three different outcomes could be expected. If listeners can only, or primarily, rely on speaking rate to make their judgements, listeners would not be expected to be able to make reliable discrimination judgements about speaking style, especially in the degraded conditions with CI simulations. If listeners use other cues in addition to speaking rate, discrimination judgements would be more difficult but still possible. Finally, if removing the unreliable cue allows listeners to shift attention to other more reliable cues, then it is also possible for discrimination judgements to be similar or better in experiment 2, compared with the overall poor performance with CI simulation in experiment 1.

EXPERIMENT 2

Methods

Participants • Twenty-eight (24 female, 4 male) NH young adults participated in experiment 2. Participants were native speakers of Dutch between the ages of 19.4 and 29.0 years (M = 22.7 years). All participants completed experiment 2 with the same hearing screening and testing conditions as experiment 1.

Materials

The same materials used in experiment 1 were also used in experiment 2. However, the utterances for both the discrimination and intelligibility tasks in experiment 2 were modified to obtain the same average speaking rate. This was achieved by using the pitch-synchronous overlap-add method (Moulines & Charpentier 1990) with the default settings (time steps of 10 msec, minimum pitch of 75 Hz, and maximum pitch of 600 Hz), similar to methods established in previous studies (Saija et al. 2014). The number of syllables and duration of each sentence was obtained automatically using a PRAAT script for detecting and counting syllables in running speech (de Jong & Wempe 2009). The average speaking rate (4.2 syllables/s) was calculated across all experimental stimuli used in both the discrimination and intelligibility tasks in experiment 1. Therefore, the duration of each utterance in both tasks was modified (either shortened or lengthened by some degree, ranging from 0.47 to 1.49 times the original duration) so that all sentences had the overall average speaking rate of 4.2 syllables/s.

Procedure • The same procedures from experiment 1 were used in experiment 2 but with the modified stimuli. In the discrimination task, trials for which no response was given within 3 sec were excluded from the analysis (approximately 4%, n = 182/4536).

Results

Speaking Style Discrimination Task • To assess whether listeners were able to make explicit judgments about speaking style without the speaking rate cue, responses for all three speaking styles were examined. In no-simulation condition, participants again gave more careful ratings for the careful read speech and more casual ratings for the casual conversation speech. Retold story speech was rated nearly equally as careful and casual. Figure 3 shows the median percent careful ratings for careful read, retold story, and casual conversation utterances.

A repeated measures ANOVA with speaking style (careful read, retold story, or casual conversation) and simulation condition (no simulation, CI-12, CI-4) as within-subject factors was carried out. The analysis revealed a significant main effect of speaking style [$F(2, 54) = 33.30$; $p < 0.001$]. The main effect of simulation condition and the interaction of simulation condition and speaking style were not significant. To examine the effect of speaking style in further detail, a series of paired comparison t tests were carried out on the responses for each speaking style. Careful read speech received significantly more careful judgements than retold story [$t(27) = 3.96$; $p < 0.001$] and casual conversation speech [$t(27) = 7.42$; $p < 0.001$], and retold story speech was rated as more careful than casual conversation speech [$t(27) = 4.68$; $p < 0.001$].

Speaking Style Intelligibility Task • To assess the intelligibility of the three different speaking styles with the same average speaking rate under the three simulation conditions, word recognition responses were examined. Figure 4 shows the word recognition accuracy in experiment 2 for careful read, retold story, and casual conversation utterances under no-simulation, CI-12, and CI-4 simulation conditions. As in experiment 1, word recognition accuracy in the no-simulation condition was near ceiling with a mean word recognition accuracy of 93.1% (SD = 5.6), while performance declined in the CI-12 and CI-4 conditions with a mean accuracy of 85.6% (SD = 6.4) and 56.6% (SD = 6.6), respectively.

A repeated measures ANOVA with speaking style (careful read, retold story, or casual conversation trials) and simulation condition (no simulation, CI-12, CI-4) as within-subject factors was carried out on word recognition accuracy, after conversion to rationalized arcsine units. The analysis revealed a significant main effect of simulation condition [$F(2, 56) = 188.57$; $p < 0.001$] and a significant main effect of speaking style [$F(2, 56) = 56.56$; $p < 0.001$], with no significant interaction between simulation condition and speaking style.

A series of paired comparison t tests on simulation condition confirmed that intelligibility in the no-simulation condition was greater than in the CI-12 condition [$t(27) = 6.08$; $p < 0.001$] and the CI-4 condition [$t(27) = 26.13$; $p < 0.001$], and intelligibility in the CI-12 condition was greater than in the CI-4 condition [$t(27) = 19.20$; $p < 0.001$]. A series of paired comparison t tests on speaking style further revealed that overall

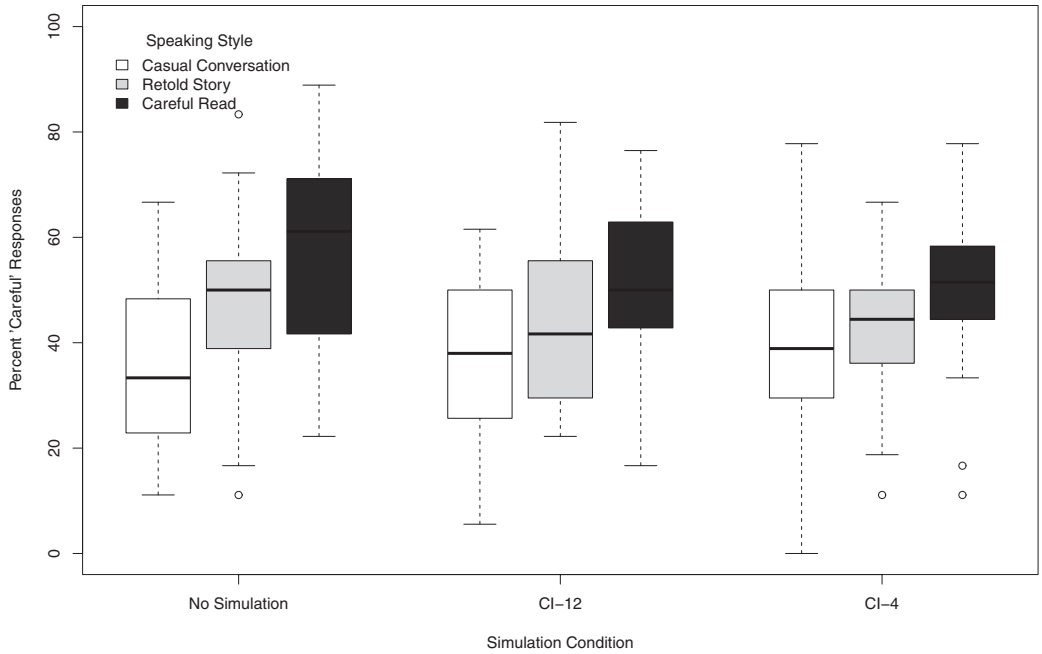

Figure 3. Percent careful ratings for careful read, retold story, and casual conversation utterances for all three simulation conditions (no simulation, CI-12, CI-4) in experiment 2. See Figure 1 for a description of the box plot design.

careful read speech was more intelligible than retold story speech [$t(27) = 9.02$; $p < 0.001$] and casual conversation speech [$t(27) = 5.84$; $p < 0.001$], and casual conversation speech was more intelligible than retold story speech [$t(27) = 2.58$; $p = 0.016$].

Discussion

Experiment 2 examined NH listeners' perception of speaking style when speaking rate cues were minimized, both with and without CI simulation. Overall, participants gave significantly more careful ratings for the careful read speech and more casual ratings for the casual conversation speech, with retold story speech rated equally often as careful and casual. Thus, the listeners were able to take advantage of other reliable cues within the utterance to categorize speaking style across the simulation conditions. Given that the different speaking styles are characterized by multiple segmental and suprasegmental cues (Tables 1–4), it is not surprising that the listeners were able to perform the task without the speaking rate cue, at least in the no-simulation condition where multiple cues were still available.

Similar to experiment 1, overall performance in experiment 2 was worse under CI simulation than the no-simulation condition, with similar performance in the CI-12 and CI-4 conditions. Although a salient (but perhaps unreliable cue) was minimized, listeners may have adopted a strategy to make use of other cues available in both the CI-12 and CI-4 conditions. As mentioned above, mean F0 and/or range differences and sound segment deletion (Tables 3 and 4) or reduction may have been reliable cues for discriminating the speaking styles in the materials used in the present study. Because these cues would have been greatly reduced in the CI-simulation conditions, the listeners may not have relied on them in experiment 1 when speaking rate, a more salient cue, was available. However, the listeners in experiment 2 may have used them in the absence of the speaking rate cue. Correlational analyses revealed that F0 mean and range were significantly related to how often an item was categorized as careful in the CI-12 condition (mean: $r = 0.23$, $p = 0.003$; range: $r = 0.26$, $p = 0.001$) but not in the no-simulation or CI-4 conditions. The overall rate of sound segment deletion (calculated as the percent of the four target sounds in Tables 3 and 4 that were deleted in each utterance) was again not significantly correlated with how often an item was categorized as careful in the CI-simulation conditions. Therefore, unlike experiment 1, F0 mean and range differences may have contributed to the perception of speaking style in the CI-12 condition in experiment 2.

Other speaking rate cues not examined in the present study, in addition to F0 mean and range and sound segment deletion, may have still been available to the listeners, such as phrase length between pauses or variability in speaking rate within fragments between pauses. While the resulting variability in speaking rate would still be available to the listener even under very degraded CI simulations, the utterances in the present study did not contain many long pauses and were not expected to display substantial differences in within-utterance speaking rate variability. Nevertheless, these differences may have contributed to speaking style perception.

Finally, the speaking style intelligibility task demonstrated that overall intelligibility was not drastically affected by the manipulation of the utterances' duration and speaking rate. As in experiment 1, performance declined as a function of amount of spectral information the listeners were receiving, but performance was not at or near floor in any condition.

GENERAL DISCUSSION

The present study investigated the perception of different real-life speaking styles by NH listeners with and without CI simulation. In both experiments, in unprocessed conditions without CI simulations, NH listeners were able to perceive reliable speaking style-specific differences among all three speaking styles and use them to make judgements about the speaking style of unfamiliar, unique utterances in both experiments.

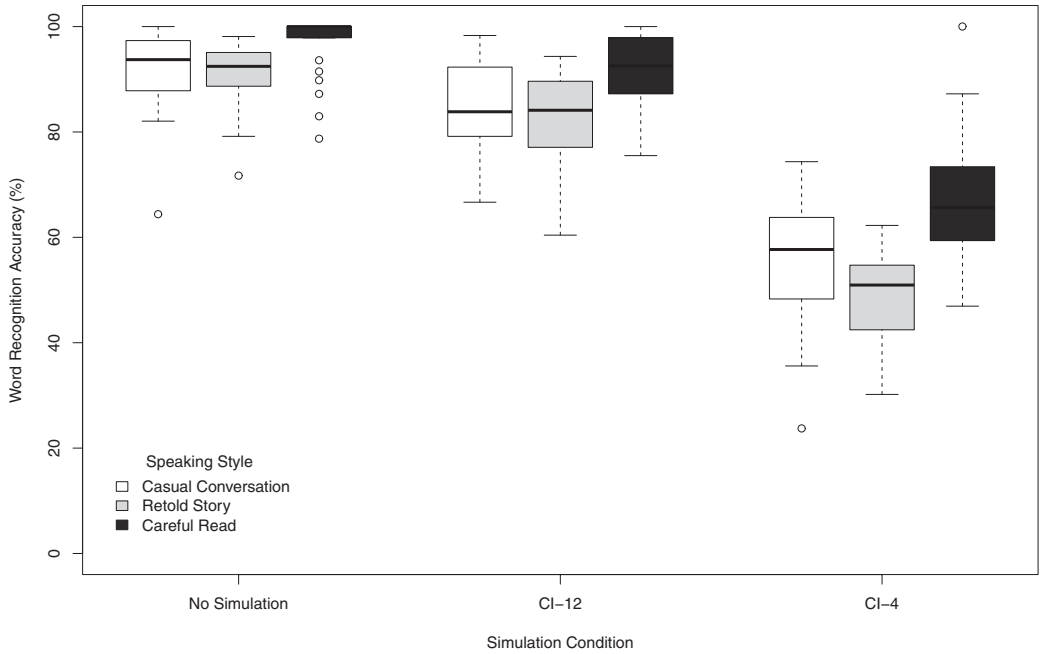

Figure 4. Word recognition accuracy in experiment 2 by speaking style (careful read, retold story, and casual conversation) and simulation condition (no simulation, CI-12, and CI-4). See Figure 1 for a description of the box plot design.

Moreover, in experiment 2, listeners were able to discriminate speaking style, even without speaking rate differences among the speaking styles. Listeners' judgements in both experiments reflected the three separate speaking style categories, with careful read speech perceived as more careful, retold story speech perceived as neither careful nor casual, and casual conversation speech perceived as more casual. These results suggest that the three speaking conditions resulted in perceptible differences in speaking style, to which listeners were sensitive, supporting prior studies demonstrating that NH listeners are able to make use of reliable cues in the speech signal to make judgements about various sources of variability, related to the talker and context (Ptacek & Sander 1966; Lass et al. 1976; Van Lancker et al. 1985a, b; Clopper & Pisoni 2004a).

Compared with the performance in unprocessed conditions without CI simulations, we predicted that speaking style discrimination would be more challenging with acoustic simulations of CI hearing. The results of both experiments partially confirmed our hypotheses that the signal degradations imposed by the CI simulation would impede successful discrimination of speaking style because many acoustic-phonetic cues indexing speaking style are carried in the spectral properties of the speech sounds. In both experiments, CI simulation clearly affected the listeners' ability to make discrimination judgements about speaking style, and perceived differences among the speaking styles were much smaller in both the CI-12 and CI-4 conditions compared with the no-simulation condition. This suggests that the spectral degradation from the CI simulation limited the amount of speaking style-specific acoustic-phonetic information that the listeners could use to make their judgements. Important cues, such as differences in F0 mean and range and the deletion and reduction of sound segments, seem not to be robustly conveyed in the simulations, hindering the discrimination of the speaking style. Thus, these findings further support previous research demonstrating that CI users and NH listeners under CI simulation have difficulty discriminating different sources of speech variability (Cleary & Pisoni 2002; Cleary et al. 2005; Massida et al. 2011; Fuller et al. 2014; Tamati et al. 2014; Gaudrain & Başkent 2015; Tamati & Pisoni 2015).

However, speaking style discrimination performance did not decline between the CI-12 and CI-4 conditions. We had predicted such decline because less reliable information about speaking style in the speech signal would be available to the listener with decreasing spectral resolution. The overall poor speaking style discrimination under CI simulation suggests that listeners may not have been relying on detailed spectral cues in either the 12- or 4-channel condition, but instead they may have been relying on an unreliable cue(s) that was available in both conditions but was not a reliable indicator of speaking style in the present set of materials. In particular, speaking rate was a likely target for discriminating speaking style degraded by CI simulation because temporal information is better maintained than spectral information in CI simulations (Fu et al. 2004; Gaudrain & Başkent 2015), and CI users show higher cue weighting of temporal cues (Winn et al. 2012; Wagner et al. 2016).

Experiment 2 allowed us to explore in more detail the perception of speaking style. Speaking rate information was altered to force listeners to use other cues, which may not have been as salient as the speaking rate cue in experiment 1. In experiment 2, although performance declined substantially under CI simulation, some differences at least between careful read speech and the others were detected, as shown in Figure 3. Because listeners appeared to be consistent in their judgements with and without CI simulation, this suggests that removal of the speaking rate cue did not entirely disrupt discrimination performance. Participants in experiment 2 may have relied on more poorly encoded but reliable information in the CI-simulation conditions. These cues may have included, for example, differences in F0 mean and range and differences in the realization of sound

segments, which are shown in Tables 3 and 4 to be consistently different among the three speaking styles used in the current task. However, this does not rule out using other potential cues for making their judgements, such as variability in speaking rate within an utterance. However, as mentioned above, because the materials did not contain many long pauses and overall average speaking rate was controlled in experiment 2, the three speaking styles likely did not vary greatly in speaking rate variability within an utterance. Additional research should be carried out with a more controlled set of stimuli to directly investigate the perception of different spectral and temporal cues that contribute to speaking style perception.

Regardless of which particular cues the listeners were using, listeners were able to discriminate the speaking styles in experiment 2, suggesting that they were able to use other cues (or sets of cues) when speaking rate was modified. These findings are broadly consistent with previous studies demonstrating the flexibility of the perceptual system to adapt and adjust to a degraded signal. NH listeners under CI simulation and CI users must adapt to a degraded signal in which important speech cues are limited (Winn et al. 2012; Fuller et al. 2014; Wagner et al. 2016; Gaudrain & Başkent 2018). CI users may be able to adopt new perceptual strategies in which they rely on acoustic-phonetic cues that are more strongly conveyed by the CI (Gaudrain et al. 2009; Winn et al. 2012; Moberly et al. 2014) or additional sources of information (phonetic, lexical, semantic, etc.) still available to them (Clarke et al. 2014). In particular, CI users and NH listeners under CI simulation have been shown to rely more on temporal information than spectral information in phonetic perception (Xu & Pfingst 2003; Xu et al. 2005; Winn et al. 2012). However, reweighting of the perceptual use of temporal and spectral cues may not result in better phonetic perception or word recognition. Moberly et al. (2014) found that individual CI users displaying perceptual strategies more similar to NH listeners, who relied more on spectral than temporal cues, showed better word recognition than CI users displaying reweighting of the temporal and spectral cues. Similarly, in the present study, while participants in experiment 1 may have relied on speaking rate because it is more strongly conveyed by the CI, this did not lead to good perception of speaking style. When the speaking rate information was modified in experiment 2, those listeners were able to use additional cues, which, although degraded, were more reliable for speaking style perception.

The ability to rely on different sources of information in the speech signal in multiple degraded conditions may also partially explain the unexpected result that discrimination performance was similar under both the 12- and 4-channel CI simulations. Above, we accounted for this finding by suggesting that listeners were relying on the same cue (or set of cues) in both CI-simulation conditions, and that this cue was conveyed equally well in both conditions. Another contributing factor could be that listeners varied in how much they attended to linguistic and nonlinguistic sources of information in the signal and were able to ignore linguistic information and focus on the nonlinguistic information in the CI-4 condition compared with the CI-12 condition. Previous studies have shown that linguistic information can interfere with the processing of nonlinguistic information. Listeners are slower to classify speakers by gender when initial phonemes vary and vice versa (Mullennix & Pisoni 1990), and listeners are also slower to classify utterances by speaking rate when phonetic information varies (Green et al. 1997). Further, familiarity with the linguistic structure of an utterance facilitates nonlinguistic judgments, such as talker voice identification (Thompson 1987; Goggin et al. 1991; Winters et al. 2008). The CI-12 condition may have provided stronger speaking style information but also irrelevant linguistic information (see section on selection of materials above). As such, listeners may have been less able to attend to the nonlinguistic cues related to speaking style. In the CI-4 condition, although it provided weaker speaking style information, that linguistic information was partially or completely masked, potentially allowing listeners to better attend to the (further degraded) speaking style information. This interpretation is consistent with previous studies with CI users, in which prelingually deaf pediatric CI users were better able to discriminate talker voices when the linguistic content produced by 2 talkers was fixed compared with when the linguistic content varied across talkers (Cleary & Pisoni 2002; Cleary et al. 2005). However, it does not account for why performance was better in the no-simulation condition, in which linguistic information was fully available. It is possible that because processing was relatively easy due to the availability of multiple redundant cues to speaking style in the no-simulation condition, listeners may have been able to show good discrimination performance despite the variable linguistic information. If this were the case, we may also see better speaking style discrimination with materials matched in linguistic content. Therefore, more research needs to be carried out to determine the most reliable cues for speaking style (and other sources of variability), how these cues are perceived in different conditions with and without CI simulation, as well as the interaction of linguistic and nonlinguistic information in CI speech perception.

The current findings suggest that real-life speech variability may be particularly challenging for CI users. As in previous studies with other sources of variability (Fuller et al. 2014; Tamati et al. 2014), the present study shows that some speaking style information may be encoded but the amount of information is likely reduced compared with the robust information NH listeners perceive and encode. As a consequence, CI users may be limited in their ability to use the available speech variability information in communication, both to make nonlinguistic judgements about different sources of variability and to facilitate the recognition of highly variable speech.

The interpretation of the results for CI users is, however, limited because questions remain about how realistic CI simulations are in capturing or predicting performance on speech perception tasks by CI users (King et al. 2012; Bhargava et al. 2014). Speech perception performance may vary depending on the nature of the CI simulation itself (Gaudrain & Başkent 2015). In addition, differences in age, language background, and experience between the young, NH listeners and CI users may influence performance. CI users display variability in speech perception performance due to variation in age at testing, age at implantation, duration of deafness before implantation, etiology of hearing loss, and experience with the CI (Blamey et al. 2013). Further, while NH listeners would have some experience dealing with noise or competing talkers, experience with CI-simulated speech is expected to be limited. In contrast, CI users would have experience listening to different speaking styles with their CIs and may have developed strategies for dealing with speech variability (Benard & Başkent 2014). Thus, to gain a more complete picture of CI perception of real-life speaking style and other sources of speech variability, it

is necessary to carry out speech perception studies with CI users with diverse hearing histories and language backgrounds using a wide range of real-life speech.

Finally, the results of the short speaking style intelligibility tasks demonstrated that the listeners had difficulty recognizing casual conversation and retold story speech under CI simulation. The observation that the careful speech was more intelligible is consistent with previous research (Janse et al. 2007; Ranbom & Conine 2007; Tucker & Warner 2007). These results may also support previous findings showing that reduced speech is especially difficult to recognize in adverse conditions, such as when context information is not available (Pickett & Pollack 1963; Ernestus et al. 2002; Janse & Ernestus 2011) or conditions are degraded, such as with hearing loss (Janse & Ernestus 2011). CI users may also have more difficulty understanding casual reduced speech, characteristic of real-life adverse conditions, compared with the careful speech, commonly used in the clinic. However, more extensive research should investigate the effects of speaking style on speech recognition by CI users because the task only included a few utterances per talker and speaking style, and the lexical and semantic content of the utterances may have differed across speaking style.

Taken together, the results of the two speaking style discrimination tasks provide a characterization of the perception of real-life speaking style with and without CI simulation. The findings contribute to our basic scientific knowledge of the perception of nonlinguistic variability in degraded conditions. Further, they provide indications for CI perception of speech variability and perception performance in real-life listening environments (cf. Koch et al. 2016). These findings, combined with future studies with CI users, will help guide the design of more sensitive and effective clinical assessment and training tools that better represent the challenges of real-life speech communication for this clinical population.

ACKNOWLEDGMENTS

We thank Britt Bosma, Wilke Bosma, Charlotte de Blecourt, Marleen Kremer, and Fergio Sismono for their assistance with the recruitment and testing of participants, scoring, and Dutch translations for this project.

Preparation of this manuscript was supported in part by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, the VIDI Grant No. 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), VENI Grant No. 275-89-035 from the Netherlands Organization for Scientific Research (NWO), 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.

The authors have no conflicts of interest to disclose.

Address for correspondence: Terrin N. Tamati, Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands. E-mail: t.n.tamati@umcg.nl.

Received October 18, 2016; accepted March 12, 2018.

REFERENCES

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine.

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). San Diego, CA: Plural Publishing, Inc.

Benard, M. R., & Başkent, D. (2014). Perceptual learning of temporally interrupted spectrally degraded speech. *J Acoust Soc Am*, *136*, 1344.

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

Blamey, P., Artieres, F., Başkent, D., et al. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiol Neurootol*, *18*, 36–47.

Bradlow, A. R., Kraus, N., & Erin, H. (2003). Speaking clearly for learning-impaired children: Sentence perception in noise. *J Speech Lang, 46*, 80–97.

Brouwer, S., Mitterer, H., Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Lang Cognitive Proc*, *27*, 539–571.

Clarke, J., Gaudrain, E., Chatterjee, M., et al. (2014). T'ain't the way you say it, it's what you say–perceptual continuity of voice and top-down restoration of speech. *Hear Res*, *315*, 80–87.

Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Ann Otol Rhinol Laryngol Suppl*, *189*, 113–118.

Cleary, M., Pisoni, D. B., Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *J Speech Lang Hear Res*, *48*, 204–223.

Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Lang Speech*, *51*(Pt 3), 175–198.

Clopper, C. G., & Pisoni, D. B. (2004a). Perceptual dialect categorization by an adult cochlear implant user: A case study. *Int Congr Ser*, *1273*, 235–238.

Clopper, C. G., & Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *J Phon*, *32*, 111–140.

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behav Res Methods*, *41*, 385–390.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology-phonetics interface*. Dissertation. Utrecht: LOT.

Ernestus, M., Baayen, H., Schreuder, R. (2002). The recognition of reduced word forms. *Brain Lang*, *81*, 162–173.

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

Faulkner, K., Tamati, T. N., Gilbert, J. L., et al. (2015). List equivalency for the clinical evaluation of speech recognition with PRESTO. *J Am Acad Audiol*, *26*, 1–13.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (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.

Fu, Q. J., Chinchilla, S., Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol*, *5*, 253–260.

Fuller, C. D., Gaudrain, E., Clarke, J. N., et al. (2014). Gender categorization is abnormal in cochlear implant users. *J Assoc Res Otolaryngol*, *15*, 1037–1048.

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

Gaudrain, E., & Başkent, D. (2018). Discrimination of Voice pitch and vocal-tract length in cochlear implant users. *Ear Hear*, *39*, 226–237.

Gaudrain, E., Li, S., Ban, V. S., Patterson, R. D. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. *Proceedings of Interspeech* 2009 (pp. 152-155), Brighton, UK.

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

Goggin, J. P., Thompson, C. P., Strube, G., et al. (1991). The role of language familiarity in voice identification. *Mem Cognit*, *19*, 448–458.

Green, K. P., Tomiak, G. R., Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Percept Psychophys*, *59*, 675–692.

Green, T., Faulkner, A., Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *J Acoust Soc Am*, *116*(4, pt 1), 2298–2310.

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

Jaekel, B. N., Newman, R. S., Goupell, M. J. (2017). Speech rate normalization and phonemic boundary perception in cochlear-implant users. *J Speech Lang Hear Res*, *60*, 1398–1416.

Janse, E., & 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., Quené, H. (2007). Coping with gradient forms of deletion and lexical ambiguity in spoken word recognition. *Lang Cognitive Proc*, *22*, 161–200.

Ji, C., Galvin, J. J. 3rd, Xu, A., et al. (2013). Effect of speaking rate on recognition of synthetic and natural speech by normal-hearing and cochlear implant listeners. *Ear Hear*, *34*, 313–323.

Ji, C., Galvin, J. J., Chang, Y. P., et al. (2014). Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants. *J Speech Lang Hear Res*, *57*, 532–554.

Johnson, K., & Mullennix, J. W. (1997). *Talker Variability in Speech Processing*. San Diego, CA: Academic Press.

King, S. E., Firszt, J. B., Reeder, R. M., et al. (2012). Evaluation of TIMIT sentence list equivalency with adult cochlear implant recipients. *J Am Acad Audiol*, *23*, 313–331.

Koch, X., Dingemanse, G., Goedegebure, A., et al. (2016). Type of speech material affects acceptable noise level test outcome. *Front Psychol*, *7*, 186.

Krause, J. C. (2001). Properties of naturally produced clear speech at normal rates and implications for intelligibility. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Laneau, J., Wouters, J., Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *J Acoust Soc Am*, *116*, 3606–3619.

Lass, N. J., Hughes, K. R., Bowyer, M. D., et al. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J Acoust Soc Am*, *59*, 675–678.

Liu, S. (2014). Clear speech perception in acoustic and electric hearing. *J Acoust Soc Am*, *116*, 2374–2383.

Massida, Z., Belin, P., James, C., et al. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hear Res*, *275*, 120–129.

Mattys, S., Davis, M. H., Bradlow, A. R., et al. (2012). Speech recognition in adverse listening conditions: A review. *Lang Cognitive Proc*, *2*, 953–978.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *J Phon, 34,* 73–103.

Moberly, A. C., Lowenstein, J. H., Tarr, E., et al. (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.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In Plack, C. J., Oxenham, A. J., Fay, R. R., Popper, A. N. (Eds.), *Pitch Perception* (pp. 234–277). New York, NY: Springer US.

Moore, B. C., & Glasberg, B. R. (1988). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *J Acoust Soc Am*, *83*, 1093–1101.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Comm*, *9*, 453–467.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Percept Psychophys*, *47*, 379–390.

Nie, K., Barco, A., Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear Hear*, *27*, 208–217.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept Psychophys*, *60*, 355–376.

Nygaard, L. C., Sommers, M. S., Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychol Sci*, *5*, 42–46.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: intelligibility differences between clear and conversational speech. *J Speech Hear Res, 29*, 434–446.

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Lang Speech*, *6*, 151–164.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9–32). San Diego, CA: Academic Press.

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *J Speech Lang Hear Res*, *9*, 273–277.

Ranbom, L., & Conine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *J Mem Lang*, *57*, 273–298.

Saija, J. D., Akyürek, E. G., Andringa, T. C., et al. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *J Assoc Res Otolaryngol*, *15*, 139–148.

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *J Phon, 39,* 96–109.

Shannon, R. V. (1992). Temporal modulation transfer functions in patients with cochlear implants. *J Acoust Soc Am*, *91*(4, pt 1), 2156–2164.

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

Tamati, T. N., & Pisoni, D. B. (2015). The perception of foreign-accented speech by cochlear implant users. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Scotland.

Tamati, T. N., Gilbert, J. L., Pisoni, D. B. (2014). Influence of early linguistic experience on regional dialect categorization by an adult cochlear implant user: A case study. *Ear Hear*, *35*, 383–386.

Thompson, C. P. (1987). A language effect in voice identification. *Appl Cogn Psychol*, *1*, 121–131.

Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbruecken.

Van Lancker, D., Kreiman, J., Emmorey, K. (1985a). Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices. *J Phon*, *13*, 19–38.

Van Lancker, D., Kreiman, J., Wickens, T. (1985b). Familiar voice recognitions: Patterns and parameters: Part II. Recognition of rate-altered voices. *J Phon*, *13*, 39–52.

Van Son, R. J. J. H., Binnenpoorte, D., van den Heuvel, H., & Pols, L. C. W. (2001). The IFAcorpus: a phonemically segmented Dutch open source speech database. *Proceedings of Eurospeech 2001*, Aalborg, Denmark.

Wagner, A. E., Toffanin, P., Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Front Psychol*, *7*, 398.

Winn, M. B., Chatterjee, M., Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. *J Acoust Soc Am*, *131*, 1465–1479.

Winters, S. J., Levi, S. V., Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *J Acoust Soc Am*, *123*, 4524–4538.

Xu, L., & Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception. *J Acoust Soc Am*, *114*(6, pt 1), 3024–3027.

Xu, L., Thompson, C. S., Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am*, *117*, 3255–3267.