ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Adult Cochlear Implant Users Versus Typical Hearing Persons: An Automatic Analysis of Acoustic–Prosodic Parameters

Tomás Arias-Vergara,[a,b,c] Anton Batliner,[c,d] Tobias Rader,[a] Daniel Polterauer,[a] Catalina Högerle,[a] Joachim Müller,[a] Juan-Rafael Orozco-Arroyave,[b,c] Elmar Nöth,[c] and Maria Schuster[a]

[a] Department of Otorhinolaryngology and Head and Neck Surgery, Ludwig Maximilians University of Munich, Germany [b] Faculty of Engineering, Universidad de Antioquia, Medellín, Colombia [c] Pattern Recognition Lab, Friedrich-Alexander University, Erlangen-Nuremberg, Germany [d] Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany

ARTICLE INFO

Article History:
Received February 28, 2021
Revision received February 21, 2022
Accepted August 17, 2022

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

https://doi.org/10.1044/2022_JSLHR-21-00116

ABSTRACT

Purpose: The aim of this study was to investigate the speech prosody of postlingually deaf cochlear implant (CI) users compared with control speakers without hearing or speech impairment.
Method: Speech recordings of 74 CI users (37 males and 37 females) and 72 age-balanced control speakers (36 males and 36 females) are considered. All participants are German native speakers and read *Der Nordwind und die Sonne* (The North Wind and the Sun), a standard text in pathological speech analysis and phonetic transcriptions. Automatic acoustic analysis is performed considering pitch, loudness, and duration features, including speech rate and rhythm.
Results: In general, duration and rhythm features differ between CI users and control speakers. CI users read slower and have a lower voiced segment ratio compared with control speakers. A lower voiced ratio goes along with a prolongation of the voiced segments' duration in male and with a prolongation of pauses in female CI users. Rhythm features in CI users have higher variability in the duration of vowels and consonants than in control speakers. The use of bilateral CIs showed no advantages concerning speech prosody features in comparison to unilateral use of CI.
Conclusions: Even after cochlear implantation and rehabilitation, the speech of postlingually deaf adults deviates from the speech of control speakers, which might be due to changed auditory feedback. We suggest considering changes in temporal aspects of speech in future rehabilitation strategies.
Supplemental Material: https://doi.org/10.23641/asha.21579171

A cochlear implant (CI) is the most suitable device in the case of severe and profound deafness when acoustic hearing amplification is not sufficient to enable reasonable speech perception. CIs can (re)establish auditory feedback and provide sufficient hearing for speech control. According to Cosetti and Roland (2010), when provided with CI in the first year of their life, even profoundly deaf children can acquire language resembling that of children without hearing loss. However, many CI users display altered speech production and understanding even after hearing rehabilitation. Horga and Liker (2006) examined segmental and suprasegmental speech properties such as vowel formant space, voiced versus voiceless differences, voice onset time, word accent production, sentence stress production, voice quality, and pronunciation quality in three groups of 10 subjects: (a) children with a CI, (b) children with profound hearing loss using hearing aids, and (c) control children without hearing loss. They found that most CI users performed worse than the control speakers but better than deaf subjects with hearing aids.

Correspondence to Tomás Arias-Vergara: tomas.arias@fau.de. Tomás Arias-Vergara is currently employed by the University Hospital Erlangen. The current work was submitted when he was a PhD student at the Universidad de Antioquia and Friedrich-Alexander University. At that time, he was also employed by the University Clinic Munich, Ludwig Maximilians University. ***Disclosure:*** *The authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.*

Adult CI users also show altered speech production such as diminished vowel precision (Neumeyer et al., 2010) and speech production quality, measured as the word recognition rate produced by adapted automatic speech recognition (Ruff et al., 2017). Concerning articulation, the directions into velocities of articulators (DIVA) model proposed by Guenther et al. (2006) explains alteration of speech due to diminished auditory feedback. Normally, speech production is constantly monitored and compared with an internal speech model in the brain that enables precise voice and articulatory control. The study presented by Guenther et al. (2004) demonstrated the effect of this internal model of speech on persons suffering hearing loss compared with control speakers. It showed that such an internal model is acquired and maintained with the use of auditory feedback. When hearing loss and therewith loss of auditory feedback occurs after speech acquisition (postlingual onset of deafness), at first, somatosensory feedback maintains precise speech production. However, in persistent lack of auditory feedback, speech production may eventually deteriorate due to a diminished precision of speech movements as a result of an economy of effort to produce speech (Perkell et al., 2007). By that, the speech production of people who suffered hearing loss after typical speech acquisition is also deteriorated (Leder & Spitzer, 1990).

Hearing loss also affects speech quality on a suprasegmental level (Öster, 1990). A Swedish study on the prosody of children provided with CI showed limitations in perception and production; prosody production was evaluated perceptually and showed very variable but, in general, poorer results than in control speakers (Lyxell et al., 2009).

Speech prosody involves the basic parameters pitch, loudness, and duration and the variability of these parameters. These parameters are related to the acoustic variables fundamental frequency (f_o), energy, length/duration, and diverse features modeling speech (Meister et al., 2009). Prosody is important for the expression of emotions and also carries linguistic information. The latter includes aspects such as word focus and sentence stress to make distinctions between questions and statements, and phrase boundary marking. For instance, stress on different syllables changes the meaning of a word (e.g., CONtract vs. conTRACT), whereas rising pitch at the end of a phrase often indicates a question. Thus, prosody is an important feature that makes speech more comprehensible.

Presuming that control of prosody production depends on sufficient auditory feedback, some aspects of prosody perception need to be taken into account considering pitch, loudness, and duration (timing). To the best of our knowledge, the auditory feedback is primarily used for controlling one's own voice. The role of auditory feedback for speech production was demonstrated for both control speakers (Martin et al., 2018) and CI users (Gautam et al., 2019). CIs limit the processing of temporal, spectral, and amplitude-related cues (Landsberger et al., 2015) due to their architecture, electrical excitation spread, and channel interactions. Some advances in speech perception could be achieved by newer coding strategies (Bolner et al., 2020; Müller et al., 2012; Punte et al., 2014; Stickney et al., 2006); however, speech perception is still restricted with CI, for example, as for the detection of f_o contours (Meister et al., 2009; Rader et al., 2017) and pure-tone discrimination (Gfeller et al., 2002) that influences speech perception (Zhang et al., 2019) and by that auditory feedback. Frequency perception plays an important role in voice control. However, for CI users, f_o detection is restricted due to the limited spectral resolution caused by the finite number of channels of the CI and the current spread at the neurons (Rader et al., 2017). Baumann and Nobbe (2006) reported reduced pitch discrimination, especially for the deep fundamental frequencies at the male and female f_o range represented by the two most apically placed electrodes in the cochlea, even in deep insertion. In some of the CI users of their study, the stimulation in the two apical electrodes was even perceived as the same frequencies. Moreover, a limited spectral resolution will likely reduce the harmonics and formant frequency information, resulting in deficits in CI users for gender categorization, speaker identification, and speech perception (Fant & Kruckenberg, 2006; Gaudrain & Başkent, 2018; Meister et al., 2009). Raising pitch might increase somatosensory feedback, and its use was reported as an attempt of male speakers suffering hearing loss to improve the perception of their own speech (Perkell et al., 1992).

Temporal resolution has also been shown to influence the perception of acoustic cues in speech, such as intonation to identify questions or statements (Chatterjee & Peng, 2008). In summary, CI users are less successful at utilizing both spectral and temporal cues compared with a control group (Winn et al., 2016). The evaluation of prosody by Holt et al. (2016) highlights the restricted perception: They examined the reaction time on prosodic cues (word stress) in sentences in prelingual deaf adult CI users and found markedly longer reaction times in comparison to control speakers. They conclude that this reflects the poorer discrimination and/or ability to use prosodic cues.

Deviations in the speech production of people with hearing loss can be represented by means of different acoustic features. For instance, f_o and sound pressure level (SPL) have been found to be higher, and the duration of utterances has been found to be longer in people with hearing loss when compared with speech recordings from control speakers (Lane & Webster, 1991; Leder et al., 1987; Mora et al., 2012; Plant & Öster, 1986; Robb & Pang-Ching, 1992). However, there is also evidence that after cochlear implantation, deaf adults can produce pitch, loudness, duration, and speech rate values that approximate those produced by control speakers in the same age range (Gautam et al., 2019; Hassan et al., 2011; Langereis et al., 1998). Regarding speech rate, Freeman and Pisoni (2017) showed that CI users had slower speech rates compared with control peers for recordings of 91 prelingually deaf CI users and 93 control speakers. Furthermore, they found that faster speech went along with higher intelligibility. This might be due to higher speech competence, better pronunciation, and by that more fluent speech (Hönig et al., 2012).

This study focuses on the speech prosody of postlingually deafened CI users compared with a group of age-balanced control speakers. Standardized speech recordings from 74 CI users and 72 control speakers are considered. Our main hypothesis is that, even after rehabilitation by cochlear implantation, the speech of the CI users exhibits deviations from that of control speakers at the suprasegmental level. Statistical analysis was performed separately for male and female speakers due to the relationship between gender and some acoustic features. For instance, female speakers have higher pitch values than the male speakers; thus, to combine male and female speakers, it would be necessary to perform pitch standardization, for example, transformation into semitones and normalization to the mean. We also compared bilateral and unilateral CI users to identify whether using one or two CIs offer an advantage for speech production. Regarding literature on speech perception, bilateral use was shown to have a positive effect (Blamey et al., 2015; Mosnier et al., 2009; van Schoonhoven et al., 2013); thus, we expect an improved speech production accordingly.

Considering former and recent studies on speech production of CI users, the acoustic analysis is usually performed manually. Due to the effort needed for this method, mostly smaller numbers of speech recordings or few participants are included in these studies. In this study, we considered automatic acoustic analysis because it enables the examination of larger cohorts and speech material. Furthermore, automatic analysis is preferred in large data sets because it can show central tendencies with a tolerance for automatic measurement errors, whereas an individualized analysis is preferred on small data sets that require careful attention for accuracy and consistency.

The parameters considered for feature extraction include pitch, loudness, duration, and rhythm. Finally, we compare and discuss the differences between CI users and the control group, between women and men, and between uni- and bilateral CI users.

Method

Considering former and recent studies on the speech production of CI users, the acoustic analysis is usually performed manually. Due to the effort needed for this method, mostly smaller numbers of speech recordings or few participants are included in these studies. In this study, we considered automatic acoustic analysis because it enables the examination of larger cohorts and speech material. Furthermore, automated analysis is preferred in large data sets because it can show central tendencies with a tolerance for measurement errors. In contrast, an individualized analysis is preferred for small data sets that require careful attention for accuracy and consistency. Automatic feature extraction can be expected to be more consistent than manual annotation (Batliner et al., 2007; Steidl et al., 2008). Any deviation from a supposed ground truth will be systematic and not impact the comparison of, for example, pitch between different groups. Note, moreover, that the notions of “gold standard” and “measurement errors” in pitch annotation are elusive: There can be quite a few borderline cases where it is not clear whether it is an octave jump, some irregular phonation, or clear nonperiodicity (Batliner et al., 1993).

We preprocess the recordings to eliminate all surplus acoustic information; then, we extract acoustic features using automatic methods and perform statistical analysis. We report the results for the whole spoken text as mean and standard deviation values.

Data

Standardized speech recordings of 74 postlingually deaf CI users (37 males and 37 females) and 72 control speakers (36 males and 36 females) are included in the experiments. For the control speakers, hearing was not tested, but none reported to wear hearing aids, having hearing problems, or having any speech issues. All recordings were done in quiet acoustic conditions; the patients were recorded at the university hospital. The speech recordings of 56 control speakers (26 males and 30 females) were performed in the hospital, and the remaining 16 control speakers (10 males and six females) are a subset of elder speakers extracted from the *PhonDat 1*[1] corpus from the Bavarian Archive for Speech Signals (BAS), reading the same story and being recorded with a 16-bit resolution at 16 kHz. All of the participants are German native speakers. The recordings performed in the clinic were captured with a Samson Meteor Microphone[2] having a frequency response between 20 Hz and 20 kHz, a signal-to-noise ratio of 96 dBA, a sampling frequency of 44.1 kHz, and a 16-bit resolution. These recordings were down-sampled to 16 kHz to match the sampling frequency of the BAS data set. All participants were asked to read the story *Der Nordwind und die Sonne* (The North Wind and the Sun), which includes six main clauses with 10 subordinate clauses in four of them (International Phonetic Association, 1999).

Table 1 shows the demographic information of the speakers considered in this study. For the CI users, the

[1]http://hdl.handle.net/11022/1009-0000-0001-D20B-6.
[2]http://www.samsontech.com/samson/products/microphones/usb-microphones/meteormic/.

Table 1. Demographic information about the speakers considered in this study.

Item	CI users		Control group	
	Male	Female	Male	Female
No. of speakers	37	37	36	36
Mean age (*SD*) in years	66 (9)	66 (9)	65 (4)	68 (7)
Unilateral CI users	25	30		
Bilateral CI users	12	7		
Mean time of CI use (*SD*) in years	4.3 (5.5)	4.1 (4.7)		

Note. CI = cochlear implant; *SD* = standard deviation.

table also includes information about the number of speakers with unilateral or bilateral implants and the years of use of CI, which is measured as the time between the surgery and the speech recording session. Only participants older than 18 years with no other known disorder of speech are included. This research was performed in compliance with the Helsinki Declaration, and the protocol was approved by the local ethics committee (Reference No. 17/516). Written informed consent was obtained from all participants. More details about the CI users considered in this study can be found in Supplemental Material S1.

Preprocessing

Noise Reduction and Resampling

Manual segmentation was performed using the audio editor software Audacity[3] in order to remove unwanted signals that may be present in the recordings. Additionally, we used a noise reduction algorithm (also with Audacity) to reduce any background noise. In summary, the algorithm works as follows: A profile of the "background noise" is extracted by computing the short-time Fourier transform (STFT) over a silence segment, for example, the segment between the beginning of the recording and the onset of speech. Then, the mean power is computed over each point of the STFT to get thresholds for each frequency bin. The STFT of the complete signal is calculated, and the sounds with energies lower than the thresholds are attenuated for noise reduction. The preprocessed audio files were down-sampled to 16 kHz and stored in the waveform audio file (wav) format.

Voice Activity Detection

An energy-based voice activity detection (VAD) algorithm was applied to automatically detect the pauses in the recordings and to compute some acoustic features considering only speech segments. The procedure of the VAD algorithm is as follows: First, the log-energy of the signal is computed from speech frames of 15 ms taken every 10 ms. Short speech frames are preferred here in order to have a better temporal resolution to capture the rapid changes in the energy of the signal. The resulting log-energy contour is normalized and smoothed by subtracting the mean and convolving the signal with a Gaussian window of 10 ms. Silence/pause segments have the lowest energy levels compared with speech sounds; thus, pause segments are detected as speech frames with log-energy levels below a certain threshold. In the case of the normalized log-energy contour, such a threshold is the median log-energy, which is computed only on the negative values. An energy-based VAD is considered in this study due to the simplicity in the implementation and to take advantage of the reduced background noise in the recordings.

Acoustic Analysis

Acoustic features are extracted automatically from the speech recordings considering prosodic features based on pitch, loudness, duration, and rhythm. The following subsections describe all features computed in this study. The scripts used in this study can be downloaded from https://bit.ly/3oactDo.

Pitch and Loudness

These parameters are analyzed by means of f_o and SPL. The f_o contour is estimated for each recording using the periodicity detector algorithm implemented in the software Praat Version 6.0.39 (Boersma, 1993; Boersma & Weenink, 2018). The f_o values (measured in Hertz) are calculated from short-time windows of 40 ms, which are extracted every 10 ms from the speech signal. Then, the mean (Mean f_o) and standard deviation (Std f_o) of the resulting f_o contour are calculated from the voiced segments, that is, speech segments where $f_o \neq 0$. The f_o features are extracted automatically using a combination of Praat and Python[4] scripts.

We are aware of the pitch doubling/halving issues that might be produced by the f_o tracking algorithm implemented in Praat. For this reason, we have included experiments investigating the role of pitch doubling/halving errors (see Supplemental Materials S2 and S3).

In the case of loudness, the sound pressure is considered to measure the amount of acoustic energy produced by a speaker (Švec & Granqvist, 2018). The SPL (measured in dB) is computed as

$$\mathrm{SPL_{dB}} = 20\log_{10}\frac{p}{p_0}. \quad (1)$$

[3]https://www.audacityteam.org/.

[4]https://www.python.org/.

where p_0 is the reference sound pressure of the air expressed in Pascal ($p_0 = 20\ \mu$ Pa) and p is the sound pressure computed as the root-mean-square value of the speech signal. In this study, p is calculated from short-time windows of 40 ms taken from the speech signals every 10 ms. The mean (Mean SPL) and standard deviation (Std SPL) of the SPL are computed for every recording. Additionally, the VAD algorithm described in the Voice Activity Detection section is used to compute the average SPL considering only the voiced segments (VAD SPL). Varying distances between speaker and microphone could introduce an intervening factor that cannot be controlled. Thus, Mean SPL values should be interpreted with caution.

Duration

Duration and ratio of speech are characterized by considering voiced sounds and pauses. Additionally, the duration of the recordings with pauses/silence (Total Length) and without it (VAD Length) is calculated. Absolute measures are considered because all speakers read the same standard text.

The voiced sounds are extracted by selecting the speech frames with $f_o \neq 0$. Then, the number of voiced segments per second (Voiced ratio) and the average duration of voiced segments (Voiced dur) are calculated from segments with a duration longer than 40 ms.

In the case of pauses, the VAD algorithm is used to locate the silence regions in order to compute the number of pauses per second (Pause ratio) and the average duration of pauses (Pause dur) within the text. The same algorithm is used to compute the duration of the recordings only with the voiced segments (VAD Length). The voiced and pause rates are measured as the number of voiced segments/pauses per second (n/s), the average duration of voiced segments/pauses are measured in milliseconds, and the duration of the recordings is measured in seconds.

Rhythm

To analyze rhythm, Ramus et al. (1999) proposed to measure the degree of vowel/consonant duration variability.[5] These features are the standard deviation of the duration of vowels (Std Voc) and consonants (Std Con). Variability of duration can also be computed as pairwise variability index (PVI) without relying on syllable division in stress-timed languages, as proposed by Grabe and Low (2002). PVI is used to measure the variability of durations in a successive vowel (PVI Voc) and consonant (PVI Con) intervals. The PVI is computed as

$$\mathrm{PVI} = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{l_k - l_{k+1}}{0.5 \times (l_k + l_{k+1})} \right|. \tag{2}$$

where l is the list with the vowel or consonant durations and m is the number of vowels or consonants. The standard deviation of the vowel/consonant duration is measured in milliseconds. The resulting fractional value is multiplied by 100 to express it as a percentage. All of the phones in the recordings are labeled automatically using the BAS CLARIN web service,[6] which performs forced alignment using an automatic speech recognition system, that is, to find the time interval for each phone given an orthographic transcription (Kisler et al., 2017). The web platform returns the transcriptions in the TextGrid format, which includes the time stamps for the words and phonemes represented in the SAMPA format.

Summary of Acoustic Features

Table 2 shows the complete set of prosodic features considered to analyze the speech production of the speakers.[7] The features are divided into four main parameters: pitch, loudness, duration, and rhythm.

Statistical Analysis

The statistical analysis is performed with the open-source package *Pingouin* (Vallat, 2018), which is written in the programming language Python 3. First, an omnibus test of normality based on D'Agostino and Pearson's (1973) study was performed on the computed acoustic parameters. From the results, we learn that not all of the features have a normal distribution; thus, we use the nonparametric Mann–Whitney U test (Mann & Whitney, 1947) to evaluate differences between CI and control speakers.

Null hypothesis testing with p values as a decisive criterion has been repeatedly criticized (see the statement by the American Statistical Association; Wasserstein & Lazar, 2016). We therefore report p values as descriptive measures and do not employ them for accepting or rejecting a null hypothesis. We adjusted the p values using Benjamini–Hochberg adjustment (Benjamini & Hochberg, 1995). Moreover, we concentrate on interpreting effect size measures, following the recommendation of the American Psychological Association (Wilkinson, 1999).

The effect size is measured by means of Cohen's d coefficient. According to Cohen (1988), the effect size can

[5]Some authors question the validity of these metrics for speech rhythm analysis (see the works of Arvaniti, 2009; White & Malisz, 2020). Note that we use the cover term *rhythm* as a sort of container for those duration-based features that model voiced/unvoiced relations. The theoretical status of "rhythm" and its different varieties is not further discussed in the text.

[6]https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface.

[7]Note that in this article, we use "Std" to denote features and "SD" to denote measured values.

Table 2. Prosodic features considered in this study.

Parameter	Feature	Description
Pitch	Mean f_o (Hz)	Average fundamental frequency
	Std f_o (Hz)	Standard deviation of fundamental frequency
Loudness	SPL (dB)	Average sound pressure level
	Std SPL (dB)	Standard deviation of sound pressure level
Duration	Total Length (s)	Total duration of the recording
	VAD Length (s)	Duration of the recording without pauses
	Voiced dur (ms)	Average duration of voiced sounds
	Pause dur (ms)	Average duration of pauses
	Voiced ratio (n/s)	Average voiced sounds produced per second
	Pause ratio (n/s)	Average pause segments produced per second
Rhythm	PVI Voc (%)	Variability of durations in successive vowels
	PVI Con (%)	Variability of durations in successive consonants
	Std Voc (ms)	Standard deviation of vowel durations
	Std Con (ms)	Standard deviation of consonant durations

be interpreted as small (d = 0.20), medium (d = 0.50), or high (d = 0.80). These thresholds should be taken with care; however, they are used here as a general guideline to provide a better understanding of the results (see Table 3).

Results

Comparison Between CI Users and Control Speakers

Tables 4 and 5 show the Mann–Whitney *U*-test results for prosodic features extracted from CI users and control speakers. The mean and standard deviation of each acoustic feature are also reported in the tables. Figure 1 shows the box plots of the prosodic features computed for CI users and control speakers.

The results confirm our main hypothesis: Postlingually deafened CI users exhibit speech deviations compared with adults with typical hearing. Such deviations were mainly observed for duration and rhythm features. For instance, CI users had considerably longer reading times than the control speakers. In the case of the male speakers, the CI users took (on average) 12 s more than the control speakers to complete the reading (Total Length [s]: 59 vs. 47, p < .001, d = 0.87). In the case of the female speakers, the CI users took 9 s more than the control group (Total Length [s]: 56 vs. 47, p < .001, d = 1.02). CI users also produced a lower Voiced ratio (n/s) than the control group (males: 1.81 vs. 2.19, p < .01, d = 0.86; females: 1.88 vs. 2.16, p < .05, d = 0.48). In male speakers, such a difference was due to a longer duration of voiced segments (Voiced dur [ms]: 384 vs. 258, p < .001, d = 0.77); in females, a lower voiced segments ratio was due to longer pauses (Pause dur [ms]: 448 vs. 411, p < .05, d = 0.51). Regarding the rhythm features, CI users (males and females) produced higher variability in the duration of vowels and consonants. Male CI users also showed higher loudness values than the control speakers; however, such differences were not observed for the females.

Table 3. Ordinal interpretation of Cohen's d effect size coefficient.

Effect size	Cohen's d interval
Small	$0.20 \leq d < 0.40$
Medium	$0.40 \leq d < 0.80$
Large	$d \geq 0.80$

Comparison Between Unilateral and Bilateral CI Users

Tables 6 and 7 show the Mann–Whitney *U*-test results for the prosodic features extracted from unilateral and bilateral CI users. Similar to the previous experiments, male and female speakers are analyzed separately. Figure 2 shows the box plots of the prosodic features computed for unilateral and bilateral CI users.

The results were inconsistent for males versus females. However, these results are likely biased due to the unbalanced and limited number of samples in the unilateral and bilateral groups. In the case of the male speakers, unilateral CI users produced considerably higher loudness values (SPL [dB]: 70 vs. 57 p < .05, d = 0.88) and shorter pauses (Pause dur [ms]: 439 vs. 515 p < .05, d = 0.98). In the case of the females, we did not find any difference after adjusting for multiple comparisons.

Discussion

In this study, an automatic analysis is performed by computing acoustic features related to nonarticulatory (prosodic) patterns in the speech of postlingually deafened CI users in comparison to control speakers. In the following,

Table 4. Mann–Whitney *U* tests between male cochlear implant (CI) users and control speakers.

Acoustic features	CI users Male (n = 37)		Control group Male (n = 36)		Mann–Whitney *U* test		Cohen's *d*
	M	*SD*	*M*	*SD*	*U*	*p*	
Mean f_o (Hz)	134	26	127	22	791	.160	0.26
Std f_o (Hz)	29	8	28	8	769	.212	0.14
SPL (dB)	69	5	67	4	909	.012	0.53
Std SPL (dB)	15	2	13	3	681	.487	0.43
Total Length (s)	59	17	47	8	1084	< .001	0.87
VAD Length (s)	49	13	39	7	1095	< .001	0.97
Voiced dur (ms)	384	214	258	78	1019	< .001	0.77
Pause dur (ms)	464	85	453	79	730	.337	0.13
Voiced ratio (n/s)	1.81	0.51	2.19	0.36	371	.001	0.86
Pause ratio (n/s)	0.37	0.10	0.39	0.07	528	.064	0.22
PVI Voc (%)	69	8	62	7	1107	< .001	1.07
PVI Con (%)	63	9	54	8	1131	< .001	1.10
Std Voc (ms)	96	35	59	21	1191	< .001	1.28
Std Con (ms)	79	45	47	32	1112	< .001	0.83

Note. The *p* values were adjusted using the Benjamini–Hochberg procedure. See Table 2 for the definition of the acoustic features. *SD* = standard deviation.

differences between control speakers and CI users are compared for each prosodic parameter, considering gender differences.

Gender Differences

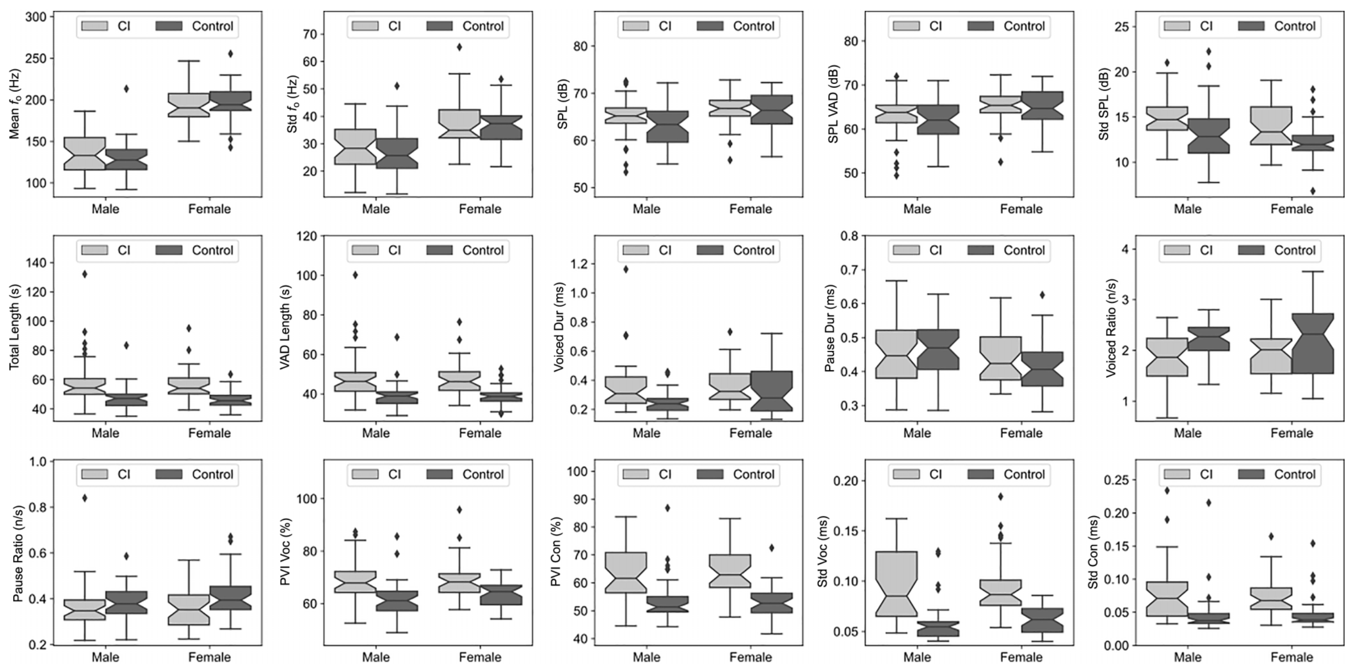
Apart from influences of hearing and auditory feedback on one's own speech, we need to keep in mind that prosody features of males and females might be different. Women generally tend toward a more "correct" pronunciation (Kreiman & Sidtis, 2011; Trudgill, 1972). This can manifest itself with a tendency toward canonical forms (less centralized vowels) as shown by Hönig et al. (2014), or with a tendency toward more isolated speech, that is, toward more and/or longer pauses. In this study, regarding gender differences in control speakers, apart from f_o features, females tend to speak louder, have shorter pauses, have higher variability of duration in successive vowels/consonants (PVI-Voc/Con), and have higher standard deviation of vowel/consonant durations (Std-Voc/Con). On rhythm features, data in the literature are rare, and one needs to keep in mind that rhythm is language dependent. Regarding other data on stress-timed languages such as German, PVI of the control speakers in this study, in general, is in the range of Grabe and Low's (2002) data for German speakers (where the gender of the

Table 5. Mann–Whitney *U* tests between female cochlear implant (CI) users and control speakers.

Acoustic features	CI users Female (n = 37)		Control group Female (n = 36)		Mann–Whitney *U* test		Cohen's *d*
	M	*SD*	*M*	*SD*	*U*	*p*	
Mean f_o (Hz)	193	26	195	23	631	.305	0.07
Std f_o (Hz)	38	9	38	8	694	.543	0.05
SPL (dB)	71	3	70	4	776	.190	0.29
Std SPL (dB)	3	2	3	2	590	.190	0.01
Total Length (s)	56	11	47	6	1057	< .001	1.02
VAD Length (s)	47	9	39	5	1102	< .001	1.15
Voiced dur (ms)	382	144	340	181	844	.060	0.25
Pause dur (ms)	448	75	411	69	878	.032	0.51
Voiced ratio (n/s)	1.88	0.46	2.16	0.68	505	.041	0.48
Pause ratio (n/s)	0.37	0.09	0.44	0.11	444	.009	0.63
PVI Voc (%)	69	7	63	5	1000	.001	0.89
PVI Con (%)	64	9	53	6	1198	< .001	1.53
Std Voc (ms)	94	31	62	13	1178	< .001	1.35
Std Con (ms)	74	30	48	24	1132	< .001	0.97

Note. The *p* values were adjusted using the Benjamini–Hochberg procedure. See Table 2 for the definition of the acoustic features. *SD* = standard deviation.

Figure 1. Box plots of the prosodic features computed for cochlear implant (CI) users and control speakers. The groups are compared considering male and female speakers separately. Mean f_o = mean of the resulting f_o contour; Std f_o = standard deviation of the resulting f_o contour; SPL = sound pressure level; VAD = voice activity detection; Std SPL = standard deviation of the SPL; Voiced Dur = average duration of voiced segments; Pause Dur = average duration of pauses; n/s = number of voiced segments/pauses per second; PVI Voc = variability of durations in successive vowels; PVI Con = variability of durations in successive consonants; Std Voc = standard deviation of the duration of vowels; Std Con = standard deviation of the duration of consonants.



speaker is not given). In accordance to our data, in Torgersen and Szakay's (2012) data of young and older English speakers with different dialects from London, men showed lower PVI values than women, whereas Szakay's (2006) data of English-speaking subjects from New Zealand revealed no gender difference for PVI. Summing up, results found in the literature on prosody regarding gender differences are inconsistent. That is partly due to the study design and might reflect language-dependent differences. The results of this study on the prosody of the

Table 6. Mann–Whitney *U* tests between unilateral and bilateral male cochlear implant users.

Acoustic features	Unilateral Male (n = 25)		Bilateral Male (n = 12)		Mann–Whitney *U* test		Cohen's *d*
	M	*SD*	*M*	*SD*	*U*	*p*	
Mean f_o (Hz)	129	25	144	23	100	.101	0.63
Std f_o (Hz)	29	8	31	9	121	.207	0.26
SPL (dB)	70	4	67	4	223	.044	0.88
Std SPL (dB)	4	3	5	3	111	.135	0.37
Total Length (s)	55	12	69	23	97	.101	0.86
VAD Length (s)	46	10	56	16	92	.087	0.86
Voiced dur (ms)	363	149	427	305	136	.681	0.30
Pause dur (ms)	439	76	515	81	68	.044	0.98
Voiced ratio (n/s)	1.89	0.50	1.64	0.50	189	.135	0.50
Pause ratio (n/s)	0.38	0.07	0.36	0.15	224	.044	0.23
PVI Voc (%)	68	8	72	6	101	.101	0.47
PVI Con (%)	62	9	67	8	90	.087	0.70
Std Voc (ms)	90	33	108	36	109	.135	0.53
Std Con (ms)	77	45	84	44	134	.331	0.16

Note. The *p* values were adjusted using the Benjamini–Hochberg procedure. See Table 2 for the definition of the acoustic features. *SD* = standard deviation.

Table 7. Mann–Whitney *U* tests between unilateral and bilateral female cochlear implant users.

Acoustic features	Unilateral Female (n = 30)		Bilateral Female (n = 7)		Mann–Whitney *U* test		Cohen's *d*
	M	*SD*	*M*	*SD*	*U*	*p*	
Mean f_o (Hz)	194	26	191	23	104	.662	0.08
Std f_o (Hz)	39	9	33	8	146	.204	0.68
SPL (dB)	71	4	71	2	109	.662	0.10
Std SPL (dB)	3	2	2	1	133	.401	0.52
Total Length (s)	56	11	56	11	119	.574	0.03
VAD Length (s)	47	9	47	8	117	.574	0.02
Voiced dur (ms)	377	144	403	139	96	.577	0.19
Pause dur (ms)	434	63	509	92	50	.128	1.08
Voiced ratio (n/s)	1.89	0.47	1.80	0.40	110	.662	0.20
Pause ratio (n/s)	0.38	0.08	0.33	0.12	157	.128	0.60
PVI Voc (%)	68	8	73	5	55	.128	0.61
PVI Con (%)	64	9	66	9	98	.662	0.19
Std Voc (ms)	92	29	102	37	81	.422	0.35
Std Con (ms)	73	28	82	36	94	.672	0.30

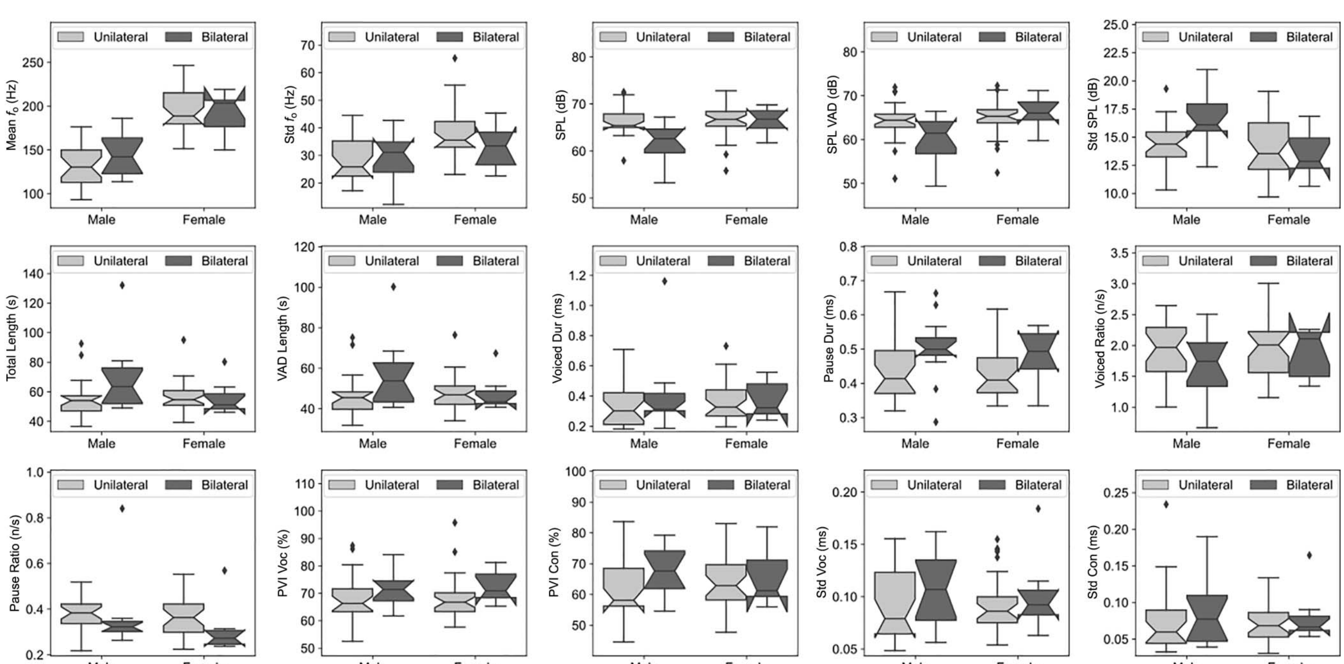
Note. The *p* values were adjusted using the Benjamini–Hochberg procedure. See Table 2 for the definition of the acoustic features. *SD* = standard deviation.

control group speakers can contribute to the discussion on this topic. Regarding prosodic characteristics in CI speakers, we refer to each parameter with respect to gender in the following, and then the prosody of uni- and bilateral CI users is compared.

Pitch and Loudness Features

Although pitch perception is known to be restricted in CI users (Gaudrain & Başkent, 2018), differences in f_o values were not found between CI users and control speakers. These results are in contrast to some found in

Figure 2. Box plots of the prosodic features computed for unilateral and bilateral cochlear implant (CI) users. The groups are compared considering male and female speakers separately. Mean f_o = mean of the resulting f_o contour; Std f_o = standard deviation of the resulting f_o contour; SPL = sound pressure level; VAD = voice activity detection; Std SPL = standard deviation of the SPL; Voiced Dur = average duration of voiced segments; Pause Dur = average duration of pauses; n/s = number of voiced segments/pauses per second; PVI Voc = variability of durations in successive vowels; PVI Con = variability of durations in successive consonants; Std Voc = standard deviation of the duration of vowels; Std Con = standard deviation of the duration of consonants.



the literature (Gautam et al., 2019; Lane & Webster, 1991; Langereis et al., 1998; Leder et al., 1987; Mora et al., 2012; Ubrig et al., 2011) where f_o of CI users tends to be higher. One reason might be that literature on pitch of CI users is mainly older and therefore represents the former state of art of hearing with CI. Due to better surgical methods and newer coding strategies, frequency representation via CI became better, enabling broader and more precise perception and auditory feedback.

All CI users included in this study lost their hearing after acquiring speech. We assume that they had access to more sources of adequate feedback before hearing loss occurred and might profit from stable feedforward processes in motor control of speech (Lane et al., 2007). Following Leder and Spitzer's (1990) data and the DIVA theory (Guenther et al., 2006), somatosensory feedback and feedforward motor control might be preserved and contribute to pitch control.

The literature on SPL of CI users is rare. In contrast to our findings, Yüksel and Gündüz (2019) found no marked differences in spectral and SPL characteristics of speech when analyzing long-term average speech spectra in postlingual CI users in comparison to control speakers. In our study, SPL and related features are similar for male and female CI users, but considerable differences are seen for male CI users in comparison with control speakers, which may reflect a higher effort in speech control in male CI users. However, further experiments involving perception are necessary to confirm this hypothesis.

Duration Features

In general, the speaking rate of people with hearing loss has been reported to be slower compared with control speakers due to a prolongation of speech segments and insertion of pauses (Osberger & McGarr, 1982; Robb & Pang-Ching, 1992). Freeman and Pisoni (2017) reported a lower speech rate in CI users. In our study, both male and female CI users exhibit a longer duration reading the text (going along with lower speech rate) and a lower voiced rate compared with the control speakers. In both male and female CI users, voiced segments and pauses are considerably longer than those produced by the control speakers.

The role of auditory feedback on reading duration was demonstrated on five postlingually deafened people after cochlear implantation by Kishon-Rabin et al. (1999). After implantation, the duration became significantly shorter on both words and sentences. In Lane et al.'s (1998) investigation of changes in duration features, they found an increased articulation rate after implantation in five of seven CI users. They conclude that "speakers use self-hearing to monitor transmission conditions and regulate speech parameters to achieve a compromise between intelligibility and effort." They refer to the theory described above that articulatory precision results from an economy of effort to produce speech (Perkell et al., 2007). In our study, the slower speech rate of CI users compared with the controls might be caused by the restricted auditory feedback that results from the degraded signal provided by the implant, reflecting the higher effort needed for these speakers. However, following Kishon-Rabin et al.'s study, no other features such as increased sound pressure support this theory.

Rhythm Features

In addition to the basic duration features, the variability of consecutive vowel and consecutive consonant durations was computed in order to model speech rhythm. According to Jang who focused on second-language learning in Koreans speaking English, these features are appropriate to represent rhythm in languages such as English and German; the variability measures matches the perceptual evaluation of rhythm (Jang, 2009; summary in Hönig, 2017).

Our results showed that CI users produced vowels and consonants with a higher duration variability compared with the control speakers. Speech rhythm has been investigated before by considering relative and absolute duration of syllables in target sentences read by people with hearing loss. In the study presented by Hood and Dixon (1969), people with hearing loss show a tendency to prolong the absolute duration of every syllable. In our study, CI users also have poorer duration control when producing different speech sounds. Moreover, in the studies reviewed by Osberger and McGarr (1982), there is evidence suggesting that people with hearing loss have difficulties controlling and coordinating the larynx and oral articulatory gestures necessary to produce, for instant, voicing contrast. In accordance to the duration features, the rhythm features PVI-Voc/Con and Std-Voc/Con are higher for male and female CI users compared with the control speakers. As for the precision of changes from voiced to unvoiced and vice versa, a previous analysis of the transitions also showed marked differences between CI and controls (Arias-Vergara et al., 2019). This may reflect the reduced auditory feedback and inconsistently increased effort to control speech.

Uni- and Bilateral Cochlear Implantation

Inconsistent results were found when comparing the acoustic features obtained from unilateral and bilateral CI users. For men, differences with effect sizes above 0.8 are found in loudness (SPL), total length, pauses and VAD length, and average pause duration. For women, differences were only seen in pause duration. In both, values of CI users with only one implant tend more toward values of the controls. These results can be explained considering the possible different representation of information on both ears (Reiss et al., 2011), which, moreover, changes over time (Reiss et al., 2007). However, considering the relatively low number of speakers in the bilateral group, conclusions are limited. Additionally, in order to analyze the benefits of two implants, it would be necessary to also evaluate the speech production of the bilateral CI users in a setting with two, one, or no implants.

Limitations

In order to have comparable speech corpora, in this study, only an identical reading text was evaluated. The motivation for choosing a reading text lies in the standardization: Reading texts usually gives stable vocabulary and length of the speech signal, enabling precise group comparisons. In future projects, also free speech should be included. However, as this shows more dependencies on the emotional state and personality, a high variation in vocabulary and narrative competence, overall words, and other features, comparability will still be a topic of concern. Future research should also include longitudinal data to monitor speech production over time. Additionally, specifics about the hearing state of the patients, side of implantation, insertion depth, active electrodes, manufacturer, filter settings of the input filters, and duration of CI usage should be taken into consideration, as well as possible influences on the acoustic parameters of speech production. In this study, prosody was evaluated referring to overall pitch, loudness, duration features and their variability, and relations of duration features. For further evaluations, phrase structure and word stress should be considered.

Conclusions

Speech production of postlingual CI users differs from that of age-matched control speakers at the suprasegmental level. In general, CI users produced markedly higher variability in the duration of vowels/consonants, took longer time for reading with longer voiced parts and pauses, and read the text with a lower voiced rate than the control speakers. The results show that even after rehabilitation by cochlear implantation, the speech of CI users deviates from typical speech. Referring to the DIVA model of auditory feedback on articulation, changes might rely on the ongoing restricted auditory feedback, yet changes are mostly seen in duration and rhythm features and not regarding pitch. The altered voiced segments' duration might reflect altered articulatory competence with changed voiced–unvoiced ratio, as shown by Osberger and McGarr (1982). The advantages of bilateral CI use are shown in many articles on speech perception (Van Schoonhoven et al., 2013) and on speech and language skills in children (Sarant et al., 2014), especially in challenging acoustic conditions such as background noise or in localization. In our study, there is no evidence for an advantage. However, our data are limited, and the advantages of using two CIs should be confirmed in a bigger cohort.

Data Availability Statement

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly; hence, supporting data are not available.

Acknowledgments

Tomás Arias-Vergara, Maria Schuster, and Elmar Nöth acknowledge the Training Network on Automatic Processing of Pathological Speech (Grant Agreement No. 766287) and Anton Batliner acknowledges EU Project sustAGE (Grant Agreement No. 826506), both funded by the Horizon 2020 program of the European Commission. Tomás Arias-Vergara is under Grants of Convocatoria Doctorado Nacional 785, financed by COLCIENCIAS.

References

Arias-Vergara, T., Orozco-Arroyave, J. R., Gollwitzer, S., Schuster, M., & Nöth, E. (2019). Consonant-to-vowel/vowel-to-consonant transitions to analyze the speech of cochlear implant users. In K. Ekštein (Ed.), *Text, speech, and dialogue. TSD 2019. Lecture notes in computer science* (Vol. 11697, pp. 299–306). Springer. https://doi.org/10.1007/978-3-030-27947-9_25

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*(1–2), 46–63. https://doi.org/10.1159/000208930

Batliner, A., Burger, S., Johne, B., & Kießling, A. (1993). MUESLI: A classification scheme for laryngealizations. In D. House & P. Touati (Eds.), *Proceedings of an ESCA workshop on prosody* (Vol. 41, pp. 176–179).

Batliner, A., Steidl, S., Schuller, B., Seppi, D., Vogt, T., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., & Aharonson, V. (2007). The impact of F0 extraction errors on the classification of prominence and emotion. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 2201–2204).

Baumann, U., & Nobbe, A. (2006). The cochlear implant electrode—Pitch function. *Hearing Research, 213*(1–2), 34–42. https://doi.org/10.1016/j.heares.2005.12.010

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blamey, P. J., Maat, B., Baskent, D., Mawman, D., Burke, E., Dillier, N., Beynon, A., Kleine-Punte, A., Govaerts, P. J., Skarzynski, P. H., Huber, A. M., Sterkers-Artières, F., Van de Heyning, P., O'Leary, S., Fraysse, B., Green, K., Sterkers, O., Venail, F., Skarzynski, H., . . . Lazard, D. S. (2015). A retrospective multicenter study comparing speech perception outcomes for bilateral implantation and bimodal rehabilitation. *Ear and Hearing, 36*(4), 408–416. https://doi.org/10.1097/AUD.0000000000000150

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences* (Vol. 17, pp. 97–110).

Boersma P., & Weenink D. (2018). *Praat: Doing phonetics by computer,* Version 6.0.32 [Computer program]. https://www.praat.org

Bolner, F., Magits, S., van Dijk, B., & Wouters, J. (2020). Precompensating for spread of excitation in a cochlear implant coding strategy. *Hearing Research, 395,* 107977. https://doi.org/10.1016/j.heares.2020.107977

Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research, 235*(1–2), 143–156. https://doi.org/10.1016/j.heares.2007.11.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cosetti, M., & Roland, J. T., Jr. (2010). Cochlear implantation in the very young child: Issues unique to the under-1 population. *Trends in Amplification, 14*(1), 46–57. https://doi.org/10.1177/1084713810370039

D'Agostino, R., & Pearson, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika, 60*(3), 613–622. https://doi.org/10.1093/biomet/60.3.613

Fant, G., & Kruckenberg, A. (2006). Individual and contextual variations of prosodic parameters. *Continuity, 4,* 5.

Freeman, V., & Pisoni, D. B. (2017). Speech rate, rate-matching, and intelligibility in early-implanted cochlear implant users. *The Journal of the Acoustical Society of America, 142*(2), 1043–1054. https://doi.org/10.1121/1.4998590

Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing, 39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

Gautam, A., Naples, J. G., & Eliades, S. J. (2019). Control of speech and voice in cochlear implant patients. *The Laryngoscope, 129*(9), 2158–2163. https://doi.org/10.1002/lary.27787

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., & Stordahl, J. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International, 3*(1), 29–53. https://doi.org/10.1179/cim.2002.3.1.29

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology, 7,* 515–546. https://doi.org/10.1515/9783110197105.515

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Guenther, F. H., Perkell, J. S., Maassen, B., Kent, R. D., & Peters, H. F. M. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 29–49). Oxford University Press.

Hassan, S. M., Malki, K. H., Mesallam, T. A., Farahat, M., Bukhari, M., & Murry, T. (2011). The effect of cochlear implantation and post-operative rehabilitation on acoustic voice analysis in post-lingual hearing impaired adults. *European Archives of Oto-Rhino-Laryngology, 268*(10), 1437–1442. https://doi.org/10.1007/s00405-011-1501-6

Holt, C. M., Demuth, K., & Yuen, I. (2016). The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants. *Ear and Hearing, 37*(4), e256–e262. https://doi.org/10.1097/AUD.0000000000000253

Hönig, F. (2017). *Automatic assessment of prosody in second language learning*. Logos.

Hönig, F., Batliner, A., Bocklet, T., Stemmer, G., Nöth, E., Schnieder, S., & Krajewski, J. (2014). Are men more sleepy than women or does it only look like—Automatic analysis of sleepy speech. In *IEEE international conference on acoustics, speech and signal processing* (pp. 995–999). https://doi.org/10.1109/ICASSP.2014.6853746

Hönig, F., Batliner, A., & Nöth, E. (2012). Automatic assessment of non-native prosody: Annotation, modelling and evaluation. In *Proceedings of the international symposium on automatic detection of errors in pronunciation training, Stockholm, Sweden* (pp. 21–30).

Hood, R. B., & Dixon, R. F. (1969). Physical characteristics of speech rhythm of deaf and normal-hearing speakers. *Journal of Communication Disorders, 2*(1), 20–28. https://doi.org/10.1016/0021-9924(69)90051-3

Horga, D., & Liker, M. (2006). Voice and pronunciation of cochlear implant speakers. *Clinical Linguistics & Phonetics, 20*(2–3), 211–217. https://doi.org/10.1080/02699200400027015

International Phonetic Association. (1999). *Handbook of the International Phonetic Alphabet: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Jang, T. Y. (2009). Rhythm metrics of spoken Korean. *Language and Linguistics, 46,* 169–185.

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., & Hildesheimer, M. (1999). The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants. *The Journal of the Acoustical Society of America, 106*(5), 2843–2857. https://doi.org/10.1121/1.428109

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language, 45,* 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies—An interdisciplinary approach to voice production and perception*. Wiley. https://doi.org/10.1002/9781444395068

Landsberger, D. M., Svrakic, S., Roland, J. T., & Svirsky, M. (2015). The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants. *Ear and Hearing, 36*(5), e207–e213. https://doi.org/10.1097/AUD.0000000000000163

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., & Zandipour, M. (2007). Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research, 50*(4), 913–927. https://doi.org/10.1044/1092-4388(2007/065)

Lane, H., Perkell, J., Wozniak, J., Manzella, J., Guiod, P., Matthies, M., MacCollin, M., & Vick, J. (1998). The effect of changes in hearing status on speech sound level and speech breathing: A study conducted with cochlear implant users and NF-2 patients. *The Journal of the Acoustical Society of America, 104*(5), 3059–3069. https://doi.org/10.1121/1.423900

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America, 89*(2), 859–866. https://doi.org/10.1121/1.1894647

Langereis, M. C., Bosnian, A. J., van Olphen, A. F., & Smoorenburg, G. F. (1998). Effect of cochlear implantation on voice fundamental frequency in postlingually deafened adults. *Audiology, 37*(4), 219–230. https://doi.org/10.3109/00206099809072976

Leder, S. B., & Spitzer, J. B. (1990). A perceptual evaluation of the speech of adventitiously deaf adult males. *Ear and Hearing, 11*(3), 169–175. https://doi.org/10.1097/00003446-199006000-00001

Leder, S. B., Spitzer, J. B., & Kirchner, J. C. (1987). Speaking fundamental frequency of postlingually profoundly deaf adult men. *The Annals of Otology, Rhinology & Laryngology, 96*(3), 322–324. https://doi.org/10.1177/000348948709600316

Lyxell, B., Wass, M., Sahlén, B., Samuelsson, C., Asker-Árnason, L., Ibertsson, T., Mäki-Torkko, E., Larsby, B., & Hällgren, M. (2009). Cognitive development, reading and prosodic skills in children with cochlear implants. *Scandinavian Journal of Psychology, 50*(5), 463–474. https://doi.org/10.1111/j.1467-9450.2009.00754.x

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., & Houde, J. F. (2018). Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience, 12,* 91. https://doi.org/10.3389/fnhum.2018.00091

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology, 48*(1), 38–48. https://doi.org/10.1080/14992020802293539

Mora, R., Crippa, B., Cervoni, E., Santomauro, V., & Guastini, L. (2012). Acoustic features of voice in patients with severe hearing loss. *Journal of Otolaryngology—Head & Neck Surgery, 31*(2), 79–88.

Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., Ambert-Dahan, E., & Ferrary, E. (2009). Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients. *Audiology and Neurotology, 14*(2), 106–114. https://doi.org/10.1159/000159121

Müller, J., Brill, S., Hagen, R., Moeltner, A., Brockmeier, S. J., Stark, T., Helbig, S., Maurer, J., Zahnert, T., Zierhofer, C., Nopp, P., & Anderson, I. (2012). Clinical trial results with the MED-EL fine structure processing coding strategy in experienced cochlear implant users. *Journal for Oto-Rhino-Laryngology and Its Related Specialties, 74*(4), 185–198. https://doi.org/10.1159/000337089

Neumeyer, V., Harrington, J., & Draxler, C. (2010). An acoustic analysis of the vowel space in young and old cochlear-implant speakers. *Clinical Linguistics & Phonetics, 24*(9), 734–741. https://doi.org/10.3109/02699206.2010.491173

Osberger, M. J., & McGarr, N. S. (1982). Speech production characteristics of the hearing impaired. *Speech and Language, 8,* 221–283. https://doi.org/10.1016/B978-0-12-608608-9.50013-9

Öster, A. (1990). The effects of prosodic and segmental deviations on intelligibility of deaf speech. *Speech Transmission Laboratory–Quarterly Progress and Status Reports, 31*(1), 65–86.

Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., & Burton, E. (2007). Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. *The Journal of the Acoustical Society of America, 121*(1), 505–518. https://doi.org/10.1121/1.2384848

Perkell, J. S., Lane, H., Svirsky, M., & Webster, J. (1992). Speech of cochlear implant patients: A longitudinal study of vowel production. *The Journal of the Acoustical Society of America, 91*(5), 2961–2978. https://doi.org/10.1121/1.402932

Plant, G., & Öster, A. M. (1986). The effects of cochlear implantation on speech production: A case study. *Speech Transmission Laboratory—Quarterly Progress and Status Reports, 27*(1), 65–86.

Punte, A. K., De Bodt, M., & Van de Heyning, P. (2014). Long-term improvement of speech perception with the fine structure processing coding strategy in cochlear implants. *Journal for Oto-Rhino-Laryngology and Its Related Specialties, 76*(1), 36–43. https://doi.org/10.1159/000360479

Rader, T., Fastl, H., & Baumann, U. (2017). Simulation of speech perception with cochlear implants. Influence of frequency and level of fundamental frequency components with electronic acoustic stimulation. *HNO, 65*(3), 237–242. https://doi.org/10.1007/s00106-016-0232-9

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Reiss, L. A. J., Lowder, M. W., Karsten, S. A., Turner, C. W., & Gantz, B. J. (2011). Effects of extreme tonotopic mismatches between bilateral cochlear implants on electric pitch perception: A case study. *Ear and Hearing, 32*(4), 536–540. https://doi.org/10.1097/AUD.0b013e31820c81b0

Reiss, L. A. J., Turner, C. W., Erenberg, S. R., & Gantz, B. J. (2007). Changes in pitch with a cochlear implant over time. *Journal for the Association for Research in Otolaryngology, 8*(2), 241–257. https://doi.org/10.1007/s10162-007-0077-8

Robb, M. P., & Pang-Ching, G. K. (1992). Relative timing characteristics of hearing-impaired speakers. *The Journal of the Acoustical Society of America, 91*(5), 2954–2960. https://doi.org/10.1121/1.402931

Ruff, S., Bocklet, T., Nöth, E., Müller, J., Hoster, E., & Schuster, M. (2017). Speech production quality of cochlear implant users with respect to duration and onset of hearing loss. *Journal of Oto-Rhino-Laryngology and Its Related Specialties, 79*(5), 282–294. https://doi.org/10.1159/000479819

Sarant, J., Harris, D., Bennet, L., & Bant, S. (2014). Bilateral versus unilateral cochlear implants in children: A study of spoken language outcomes. *Ear and Hearing, 35*(4), 396–409. https://doi.org/10.1097/AUD.0000000000000022

Steidl, S., Batliner, A., Nöth, E., & Hornegger, J. (2008). Quantification of segmentation and F0 errors and their effect on emotion recognition. In *Proceedings of International conference on text, speech and dialogue* (pp. 525–534). https://doi.org/10.1007/978-3-540-87391-4_67

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., & Opie, J. M. (2006). Effects of electrode design and configuration on channel interactions. *Hearing Research, 211*(1–2), 33–45. https://doi.org/10.1016/j.heares.2005.08.008

Švec, J. G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research, 61*(3), 441–461. https://doi.org/10.1044/2017_JSLHR-S-17-0095

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In *Proceedings of the 11th Australasian international conference on speech science & technology, University of Auckland* (pp. 421–426).

Torgersen, E. N., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua, 122*(7), 822–840. https://doi.org/10.1016/j.lingua.2012.01.004

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society, 1*(2), 179–195. https://doi.org/10.1017/S0047404500000488

Ubrig, M. T., Goffi-Gomez, M. V. S., Weber, R., Menezes, M. H. M., Nemr, N. K., Tsuji, D. H., & Tsuji, R. K. (2011). Voice analysis of postlingually deaf adults pre- and post-cochlear implantation. *Journal of Voice, 25*(6), 692–699. https://doi.org/10.1016/j.jvoice.2010.07.001

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software, 3*(31), 1026. https://doi.org/10.21105/joss.01026

Van Schoonhoven, J., Sparreboom, M., van Zanten, B. G. A., Scholten, R. J. P. M., Mylanus, E. A. M., Dreschler, W. A., Grolman, W., & Maat, B. (2013). The effectiveness of bilateral cochlear implants for severe-to-profound deafness in adults: A systematic review. *Otology & Neurotology, 34*(2), 190–198. https://doi.org/10.1097/MAO.0b013e318278506d

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

White, L., & Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 166–179). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.10

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Winn, M. B., Won, J. H., & Moon, I. J. (2016). Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization. *Ear and Hearing, 37*(6), e377–e390. https://doi.org/10.1097/AUD.0000000000000328

Yüksel, M., & Gündüz, B. (2019). Long-term average speech spectra of postlingual cochlear implant users. *Journal of Voice, 33*(2), 255.e19–255.e25. https://doi.org/10.1016/j.jvoice.2017.10.013

Zhang, F., Underwood, G., McGuire, K., Liang, C., Moore, D. R., & Fu, Q. J. (2019). Frequency change detection and speech perception in cochlear implant users. *Hearing Research, 379,* 12–20. https://doi.org/10.1016/j.heares.2019.04.007