Original Article

Approximations to the Voice of a Cochlear Implant: Explorations With Single-Sided Deaf Listeners

Trends in Hearing
Volume 24: 1–12
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216520920079
journals.sagepub.com/home/tia
SAGE

Michael F. Dorman[1], Sarah Cook Natale[1], Leslie Baxter[2], Daniel M. Zeitler[3], Matthew L. Carlson[4], Artur Lorens[5], Henryk Skarzynski[5], Jeroen P. M. Peters[6], Jennifer H. Torres[7], and Jack H. Noble[8]

Abstract

Fourteen single-sided deaf listeners fit with an MED-EL cochlear implant (CI) judged the similarity of clean signals presented to their CI and modified signals presented to their normal-hearing ear. The signals to the normal-hearing ear were created by (a) filtering, (b) spectral smearing, (c) changing overall fundamental frequency (F0), (d) F0 contour flattening, (e) changing formant frequencies, (f) altering resonances and ring times to create a metallic sound quality, (g) using a noise vocoder, or (h) using a sine vocoder. The operations could be used singly or in any combination. On a scale of 1 to 10 where 10 was a complete match to the sound of the CI, the mean match score was 8.8. Over half of the matches were 9.0 or higher. The most common alterations to a clean signal were band-pass or low-pass filtering, spectral peak smearing, and F0 contour flattening. On average, 3.4 operations were used to create a match. Upshifts in formant frequencies were implemented most often for electrode insertion angles less than approximately 500°. A relatively small set of operations can produce signals that approximate the sound of the MED-EL CI. There are large individual differences in the combination of operations needed. The sound files in Supplemental Material approximate the sound of the MED-EL CI for patients fit with 28-mm electrode arrays.

Keywords

cochlear implant, single-sided deafness, sound quality

Received 13 November 2019; Revised 23 March 2020; accepted 27 March 2020

This study (a) reviews a multiyear and multicenter effort to create approximations to the sound of a cochlear implant (CI) and (b) describes the approximations created in a recent study using English-speaking patients. The research was made possible by the innovation of fitting single-sided deaf (SSD) patients with a CI (Van de Heyning et al., 2008). This intervention has been used in Europe for over a decade (e.g., Arndt et al., 2011) and has recently been approved in the United States for the MED-EL CI.

SSD-CI patients can judge the similarity of a clean signal presented to their CI ear and candidate, CI-like signals presented to their normal-hearing (NH) ear. Our goal was to create signals for the NH ear which were close approximations, or in the best case an exact match, to the sound of a CI.

[1]Speech and Hearing Science, College of Health Solutions, Arizona State University
[2]Department of Clinical Neuropsychology, Mayo Clinic Arizona
[3]Department of Otolaryngology/Head-Neck Surgery, Virginia Mason Medical Center
[4] Department of Otolaryngology-Head and Neck Surgery, Mayo Clinic, Rochester, Minnesota, United States
[5]World Hearing Centre, Institute of Physiology and Pathology of Hearing, Warsaw, Poland
[6]Department of Otorhinolaryngology and Head & Neck Surgery, University Medical Center Utrecht
[7]Denver Ear Associates, Englewood, Colorado, United States
[8]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, Tennessee, United States

Corresponding Author:
Michael Dorman, Speech and Hearing Science, College of Health Solutions, Arizona State University, Tempe, AZ 85287, United States.
Email: mdorman@asu.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Value of the Data

When a patient first qualifies for a CI, she/he commonly asks: "How well will I be able to hear?" and "What will it sound like?" Researchers have collected large sets of data to answer the first question for adults (e.g., Blamey et al., 2013) and children (e.g., Manrique et al., 2019; Tobey et al., 2013). However, there is very little information to answer the second question. This is surprising given that CIs are over a half-century old. To compound the problem, the noise-vocoder "simulations" of CIs, found on the internet, are not good representations of CI sound quality. Thus, when we started this project, the simulations on the internet were misleading and there were no other data for patients to review.

A second reason for collecting data of this type was to provide new input to designers of CIs. For the last half century, the principal input has been percentage correct scores on tests of speech understanding. This is reasonable, as the goal has been to enable patients to understand speech. However, there have been no fundamental breakthroughs in CI signal coding for many years and speech understanding scores have stagnated (Wilson et al., 2016). A new kind of input—about the sound quality or voice of CIs—may foster the development of new and better coding strategies.

Preliminary Experiments

In 2006, the first author's research group tested a CI patient whose pure tone thresholds in the ear opposite the CI were relatively flat at 65 to 75 dB HL from 250 Hz to 8 kHz. The wide range of audible frequencies in this ear (a) allowed the creation of a wide-band, electric frequency-to-place-of-stimulation map for the patient (Dorman et al., 2007) and (b) enabled our initial attempt to match the sound quality of a CI. The patient's verbal description suggested that the CI was missing both low and high frequencies. A band-pass filtered, sine-vocoder signal produced the best match. However, the hearing loss in the better ear confounded an interpretation of this result.

In 2014, the first author's research group began testing SSD-CI patients with NH thresholds in the ear contralateral to their CI. The first patient had been followed for several years in a study of cortical plasticity (Sharma et al., 2016). The patient was born with NH bilaterally and began to lose hearing in one ear at age 5. By age 10, her hearing was sufficiently poor in that ear that she was fit with an MED-EL CI. Auditory thresholds in her better ear were less than or equal to 10 dB HL from .25 to 8 kHz. Before surgery, cortical-evoked potentials revealed cortical reorganization induced by the years of unilateral stimulus deprivation. Several years after surgery, the cortical markers of deprivation and reorganization had disappeared and the patient presented with normal or near normal cortical responses.

Approximately 3 years following device activation, the patient was asked to match the sound quality of her implant. The match was to a signal that was band pass filtered between 400 Hz and 1 kHz and had been flanged—a delay and add back operation that produces a slight amplitude and F0 shift over time. The match, and her confidence in that match, is shown in Supplemental Material 1. As shown in the video, the patient dismissed both sine and noise vocoders peremptorily. Critically, the patient commented that her CI sounded like someone was talking from behind a door or was talking with a hand over her mouth. This observation was consistent with her match to a band-pass filtered signal and was consistent with the report of the patient tested in 2006.

In 2015, the second author tested in English a bilingual group of SSD-CI patients in the Netherlands. The patients used Cochlear Corporation devices. Five of the six patients rated their matches as a 7 or 8 on a scale from 1 to 10 (a complete match). One patient rated his match as 9.8. Audiometric thresholds in this patient's better hearing ear were within the range of normal from .25 kHz to 4 kHz. The match was a combination of a 20-channel noise vocoder and a 2 kHz low-pass filter (see Supplemental Material 2). Thus, filtering, and a muffled percept, played a part in obtaining a close match to CI sound quality. However, our inability to achieve an excellent match to CI sound quality for most patients prompted an exploration of more stimulus dimensions.

In 2016, the second author tested nine Polish SSD-CI patients fit with MED-EL devices. For eight of the nine patients, match scores were either 9 or 9.5 of 10 (see Supplemental Material 3). To make a match, most patients needed multiple operations on a clean signal. Seven of the nine patients needed either low-pass or band-pass filtering, six of the nine needed spectral peak broadening (Baer & Moore, 1993), five of the nine needed either an increase in F0 or overall frequency spectrum, four of the nine needed a flange component, and two of nine needed a noise vocoder component. The good matches to CI sound quality provided by the signals used in this study lead to the use of similar signals in this study.

Vocoders as an Acoustic Model of the Sound Quality of a CI

Noise vocoders were introduced to a large audience by Shannon et al. in 1995. Shannon's critical observation was that a high level of speech understanding could be achieved with only a small number of channels. Although Shannon et al. (1995) made no claim that a CI *sounded* like a noise vocoder, given the vocoder-like

signal processing of CIs (Loizou, 1998, 2006) and the absence of any other tool, vocoders were a good first choice to begin the study of CI sound quality.

Svirsky et al. (2013) studied a single patient fit with a Cochlear Corporation Freedom CI and the advanced combination encoder (AEC) processing strategy and reported a "very similar" match to CI sound quality using a noise vocoder with an upshifted set of input filters.We return to the finding of upshifted input filters later in this study. Although a "very similar" match was obtained, Svirsky et al. cautioned against extrapolating results of acoustic model studies to actual CI users.

Dorman et al. (2017) studied nine SSD-CI patients (seven MED-EL, one Cochlear Corporation, and one Advanced Bionics) and found that vocoders, in general, provided a very poor match to CI sound quality. The mean similarity scores for noise and sine vocoders to CIs were 1.7 and 2.9, respectively, on a 10-point scale. In the previously described experiment with speakers of Polish, mean similarity scores for noise and sine vocoders were between 2.0 and 3.0.

Peters et al. (2018) studied 10 SSD-CI patients fit with Cochlear Corporation CIs (type 422 with CP 910 or 920 processors) using noise and sine vocoders as input to the NH ear.Two of the three vocoders differed in configuration from those used in Dorman et al. (2017). On a 1 to 10 scale, the average score for similarity between the vocoders and the CI was 6.8 for speech stimuli and 6.3 for music stimuli.

Most recently, Karoui et al. (2019) assessed the similarity of three vocoders (sine, noise, and pulse-spreading harmonic complex) and three output maps (no upshift in frequency, two equivalent rectangular bandwidth [ERB] upshift, and four ERB upshift) to the sound of Cochlear Corporation CIs. The pulse-spreading harmonic complex was judged more similar to a CI than either sine or noise and either no shift or a two ERB shift were judged more similar than the four ERB shift. The mean similarity score (to the sound of the CI) was 4.7, 5.1, and 5.8 for the pulse-spreading harmonic complex, noise, and sine vocoders, respectively, where 1 = *a close match* and 7 = *a poor match*. This study, our earlier studies employing vocoders, and the study of Peters et al. (2018) suggest that it would be reasonable to explore signals other than those generated by vocoders in the search for matches to the sound of a CI.

The Present Study

Motivated by the results of the preliminary experiment with Polish SSD-CI patients, in the present experiment, 14 English-speaking SSD-CI patients were asked to judge the similarity of signals delivered to their CI and modified, natural speech signals delivered to their NH ear. Other outcomes from this project, on sound source localization, on speech understanding and on the mechanism underlying "Mickey Mouse™" voice quality, are described in Dorman et al. (2015, 2017), Dorman, Natale, Zeitler, et al. (2019a), Dorman, Natale, Baxter, et al. (2019b), Zeitler et al. (2015), and Zeitler and Dorman (2019).

Methods

Participants

The listeners were 13 SSD-CI patients implanted with a 12 electrode, MED-EL 28-mm Flex Soft array and 1 patient implanted with a 31.5-mm Standard array. Five of the CIs were implanted in the left ear and nine in the right ear. For six listeners, all electrodes were activated in the clinical program; for five listeners, Electrode 12 (the most basal electrode) was deactivated; for two listeners, Electrodes 12 and 11 were deactivated and for one listener, Electrodes 12, 11, and 10 were deactivated. All listeners were fit with one of the fine structure (FS) family of signal processors (Riss et al., 2014). For each patient, an audiogram, for the ear contralateral to the implant, was provided by the referring audiologist. Demographic, clinical, and baseline data are summarized in Table 1.

High-Resolution Computed Tomography

All listeners received a postimplant, high-resolution, computed tomography scan at St. Joseph's Hospital and Medical Center in Phoenix, Arizona. The data from those scans were evaluated at Vanderbilt University using the Oto-Pilot software program (Noble et al., 2013) which outputs angular insertion depth for each electrode and frequency values for the nearest tissue in the spiral ganglion (SG) to each electrode. Figure 1 shows, for the patients in this sample, the mean SG place frequency for each electrode and the center frequency (CF) of the input filter for each electrode.

The offset in frequency between filter CF and SG place frequency is in keeping with previous reports, for example, Landsberger et al. (2015) and, presumably, is the basis for some listeners needing F0 or spectral upshifts to match the sound of their CI. The data from Oto-Pilot were made available to the researchers *after* the behavioral data from the patients had been collected. Thus, the author (S.C.N.) conducting the matching experiments did not know the SG location of the electrode contacts when conducting the experiments.

Test Signals

Three sentences from the City University of New York corpus were used for testing: (a) Do you like camping?

Table 1. Demographic, Clinical, and Baseline Data for Patients.

Patient	Gender	Processing strategy	AzBio quiet score (%)	Age	Etiology	Duration of deafness (years)	Duration of CI use (years)	Number of active electrodes	Four-frequency PTA .5, 1, 2, 4 kHz
1	M	FS4	45	65	Autoimmune disease	6.3	2	12	11
2	F	FS4	44	42	Unknown	13	2.2	11	6
3	F	FS4	88	49	Trauma	4.5	1.6	12	14
4	M	FS4p	84	47	Unknown	0.5	1.9	12	9
5	F	FS4p	100	17	Viral Infection	1.6	2.3	11	6
6	F	FS4p	100	41	Unknown	0.9	3.8	12	1
7	F	FS4	84	74	Unknown	4.8	2.3	12	10
8	F	FSP	95 (Pediatric)	12	Unknown	4.1	2.7	11	1
9	F	FS4p	52	65	Unknown	19	2.1	11	15
10	F	FS4	70	61	Autoimmune disease	1.9	4.4	10	9
11	M	FS4	71	64	Unknown	3.9	2.9	11	10
12	M	FS4p	82	70	Unknown	18	3	10	12
13	M	FS4	76	48	Unknown	5.9	0.2	12	19
14	F	FSP	46	57	Unknown	8	1.8	9	16

Note. FS = fine structure; CI = cochlear implant; PTA = pure tone average.

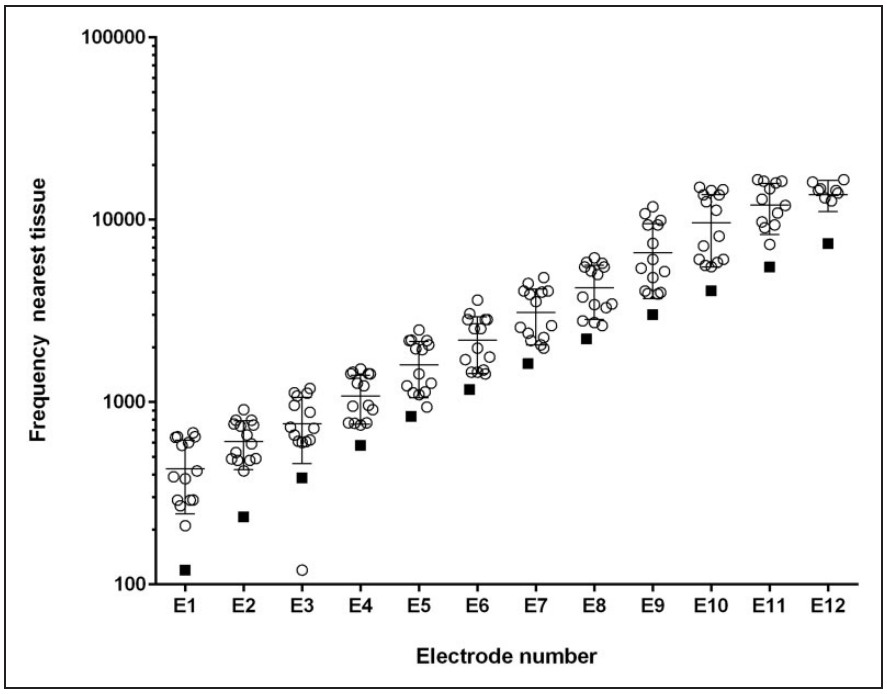

Figure 1. Estimated Spiral Ganglion Frequency as Function of Electrode Number. Open circles = individual patients. Filled squares = CF of filter. The mean and standard deviation are superimposed on the individual data points at each electrode location.

(b) The sun is finally shining, and (c) I like to play tennis. The sentences were first synthesized using the STRAIGHT (Kawahara et al., 2001) algorithm so that other manipulations, for example, F0 and formant shifting, could be implemented. These sentences were chosen because synthesized version and the natural version were nearly indistinguishable. One sentence was used for each subject's test session.

Custom-built software allowed changes in the following acoustic characteristics of clean speech signals.

- Signals could be low, high, and band pass filtered using sixth-order Butterworth filters with variable corner frequencies. Low-pass and band-pass filtering produce a muffled sound quality. Band-passed signals commonly sound as if they are farther away than

wide-band signals. Filtering can also create a "tinny" sound quality and the sound quality of a "transistor AM radio"—common descriptions of CI sound quality.

- Spectral peaks could be broadened and spectral peak-to-valley differences reduced in a simulation of the effects of poor frequency selectivity (the algorithm was modeled after Baer & Moore, 1993). For a synthetic-vowel test signal, with smear = 0, the F1 spectral peak-to-valley amplitude difference was 23.9 dB; with smear = 5 (moderate smear), the difference was 17.1 dB; and with smear = 10 (maximum smear), the difference was 11.8 dB. At high levels of broadening, a low level of a static-like sound was inadvertently introduced. Because the amplitude of spectral peaks for voiced sounds falls with frequency, at high degrees of broadening signals have a low-pass characteristic and sound muffled.
- Formant frequencies could be altered over the range −300 to +1000 Hz. In our implementation using STRAIGHT, the difference in frequency between formants was maintained and the whole spectrum was shifted up or down in frequency linearly. Research from our laboratory using listeners with normal auditory thresholds indicates that a 300 to 400 Hz or more upshift in spectrum, without a change in F0, produces a voice quality similar to the Munchkin characters in the movie, *The Wizard of Oz* (1930; Dorman, Natale, Zeitler, et al., 2019).
- The mean F0 could be increased or decreased and the F0 contour could be flattened in steps from 100% to 0% of the normal extent. The F0 algorithm kept the mean F0 the same as the original file and altered the end points. A completely flattened F0 contour commonly elicits the percept of a robotic talker. A large increase in F0 per se elicits a Mickey Mouse-like percept.
- Speech signals could be made to sound more or less metallic by altering resonances and ring times. A filter was constructed using a bank of sharp, inharmonically related resonances in combination with a band-pass filter. The resonant frequencies were $f =$ (442,578, 646, 782, 918, 1054, 1257, 1529, 1801, 2141, 2549, 3025, 3568, 4248 Hz). The band-pass filter had the following characteristics: passband = 442 to 4248 Hz; passband-to-stopband transition bandwidth = 100 Hz; stopband attenuation = 40 dB.
- A slight frequency and amplitude shift over time was implemented by creating a signal that was 0.01% longer than the original signal and then combining the two. Perceptually the combined signal sounded slightly comb filtered. This operation was suggested to us by a patient who played guitar and who said his implant sounded "flanged."
- Noise and sine vocoders could be implemented with 4 to 12 channels (vocoder parameters are described in Dorman et al., 2017). A noise vocoder has a hissy quality, and a sine vocoder has an electronic "whine." Neither evokes a strong impression of pitch.

Procedure

The matching procedure was controlled by author S.C. N. who operated a software mixing console with sliders for the dimensions listed earlier. The matching procedure is shown in the video in Supplemental Material 4 (other videos of the matching procedure, for patients fit with a different CI, are found in Dorman, Natale, Baxter, et al., 2019; Dorman, Natale, Zeitler, et al., 2019). Signals were delivered to the CI via a direct connect cable, and signals were delivered to the NH ear via an insert receiver (ER3-A). A clean signal was delivered to the CI first and then to the NH ear. The patient was asked how the signal to the NH ear should be changed to sound like the signal to the CI ear. A list of 64 audio terms was made available to patients who needed help with descriptions of sound quality. The list is found in Appendix and included terms like emphasized bass, blurred, Chipmunk-like, clear, computer-like, crisp, Darth Vader-like, distorted, far-away, grainy, robotic, scratchy, unclear, high pitched, low pitched, flat/monotone, rounder/softer, muffled, far away, and in a tube.

If a patient said, for example, that the signal to the CI ear was high pitched relative to the signal presented to the NH ear, then the experimenter presented a signal to the NH ear with a higher F0 and asked if this signal was closer to the sound of the CI. Then, the experimenter presented a signal with an upshifted formant pattern and asked if this was closer to the sound of the CI. The patient indicated which type of upshift was necessary or whether both were necessary. The experimenter then manipulated F0 and formant upshift values until the patient said that the signal to the NH ear was similar in pitch to that of the CI.

The patient was then asked what else needed to be changed. If the answer, for example, was that the CI signal sounded muffled or far away, then signals that were low-pass filtered or band-pass filtered were presented to the NH ear. This process, asking the patient what needed to be changed and then altering the signal, continued until the patient said that the match was very close. If there was an acoustic dimension that had not been manipulated, then signals were modified using that dimension and played to the patients in order to make sure that the match could not be improved. At this point, the patient was asked to rate the similarity of the signal presented to the NH ear relative to that of the CI on a 10-point scale with 10 being a complete match. All procedures were approved by the Arizona State University Review Board for the Protection of Human Subjects (institutional review board).

Results

The match scores for all patients are shown in Table 2 and Figure 2.

The range of scores was 7 to 10 with a mean score of 8.8. Over half of the scores were 9.0 or higher. One patient ranked her match as a 10. Sound files for all matches are found in Supplemental Materials 4.

The operations that contributed to the match and the number of occasions they were employed are shown in Figure 3.

The mean number of operations was 3.4. A filtering operation was the most common modification—11 of 14 patients needed this operation to match the sound quality of their CI. Spectral peak broadening was the next most common operation, followed, in order, by F0 contour flattening, F0 upshift, spectral upshift, flange, metallic, noise vocoding, spectral downshift, and sine vocoding. In the cases in which vocoding was used, the signals were layered or mixed with nonvocoded signals in order to produce either a noise component or a tonal component.

Pitch contour flattening was not implemented for 4 of the 14 patients. These patients were tested before this operation became available. Of the 10 remaining patients, 7 needed some degree of F0 contour flattening to improve their match to the sound of the CI.

Discussion

Our aim was to create signals that were close approximations to the sound of a CI. The data shown in Figure 2 and Table 2 suggest that we were generally successful in achieving this aim.

Operations Necessary to Match CI Sound Quality

The two most common operations used to create a matching signal were (a) band-pass or low-pass filtering and (b) spectral peak broadening. The common perceptual consequence of these operations was a more muffled percept or less muffled percept depending on the filter settings and degree of broadening.

That CI signals should sound muffled is not surprising given the many abnormalities engendered by electrical stimulation. Poor spectral resolution (e.g., Gifford et al., 2018; Zeng et al., 2014) and abnormalities in temporal waveforms (Wilson et al., 1997) are likely candidates for mechanisms underlying this effect. Because these two abnormalities likely occur for all patients, the muffled

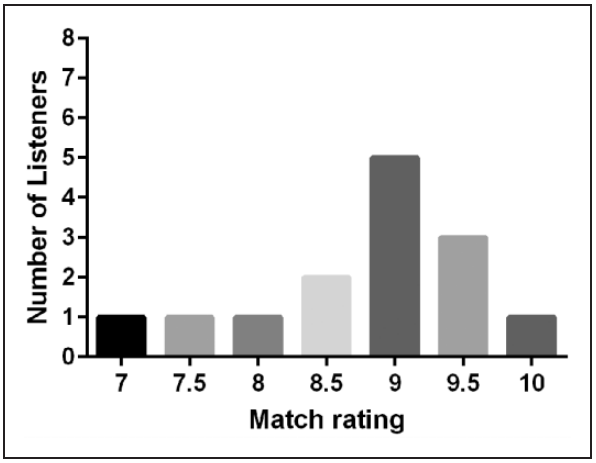

Figure 2. Ratings for Similarity of Simulation to CI Sound Quality.

Table 2. Outcome Data for Patients in Study.

Patient	Insertion angle	SG freq. E1 (Hz)	Change in F0 (Hz)	Change in formant freq. (Hz)	Filtering	Broad.	F0 contour	Flange	VC	Metallic	Match rating
1	592	270	−20	−300	LP .7 kHz	Mod	NI			Yes	8.5
2	586	210	20	0	LP 3 kHz		0.5	Yes			9.5
3	584	292	30	0	LP 2 kHz		0.5		Noise VC 35%		9
4	584	292	0	0	BP .2–1.7 kHz	Max	0.5				9
5	581	300	−10	0		Min					10
6	576	300	0	0	LP .5kHz		NI				9
7	525	390	0	0	BP .4–4 kHz	Min	NI	Yes			9
8	500	440	5	0	LP .7 kHz						9.5
9	474	534	20	300	BP .4–3 kHz	Max	0.75				9
10	460	580	0	100	HP .4 kHz	Min	0.25	Yes			8
11	455	600	−10	0		Mod	0.75				7.5
12	426	645	60	300			0.25				7
13	415	642	10	0	BP .7–1 kHz		NI		Sine VC −6 dB		9.5
14	394	680	−20	200	BP .7–1.5 kHz	Max				Yes	8.5

Note. SG = spiral ganglion; LP = low pass; HP = high pass; BP = band pass; freq. = frequency; broad. = broadening; mod = moderate; min = minimal; max = maximum; VC = vocoder; F0 = fundamental frequency of voice; E1 = electrode 1 or most apical electrode; NI = not implemented.

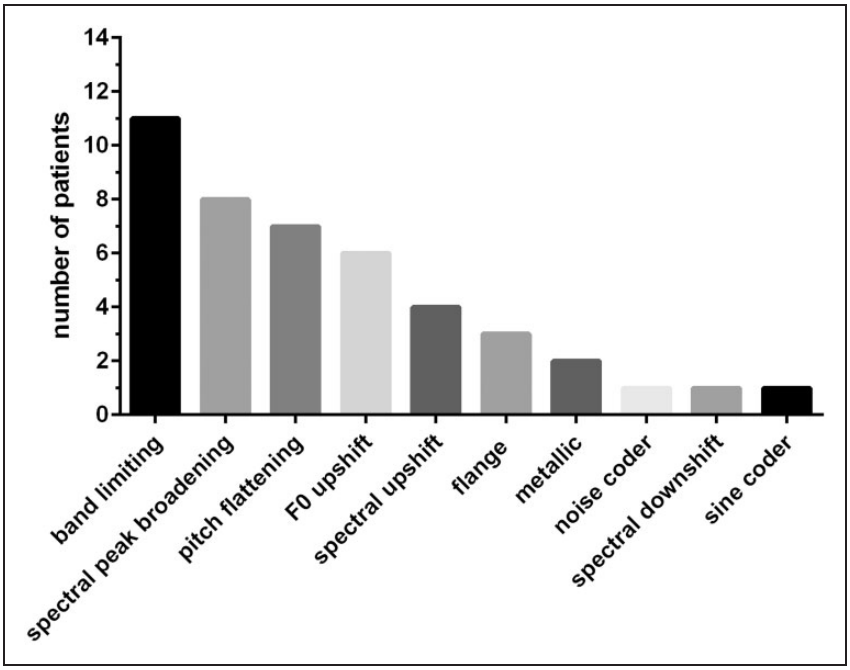

Figure 3. Operations Needed to Make Matches to CI Sound Quality.

percept should not be unique to the patients and signal processors/electrode arrays tested in this study.

A reduced F0 contour was the third most common operation needed to match the sound quality of the CIs. This outcome probably arises from the poor frequency resolution of CI patients (e.g., Zeng et al., 2014).

F0 changes and spectral upshifts were the next most common operations to match the sound quality of the CIs. As shown in Figure 1, all input signals were delivered to electrodes that, on average, were near much higher SG frequencies than the input filter CFs. However, there were large individual differences in the distance between filter CF and SG frequency. These differences were notable given that all electrode arrays, with one exception, were of the same length. As shown in Table 2, the patients with relatively shallow insertions, for example, the most apical electrode at less than 500° insertion angle (or higher than approximately 450 Hz SG place frequency), commonly needed an upshift in spectrum to match the sound quality of their CI. These data suggest that devices with relatively short electrode arrays may have a different sound quality than the device and electrode arrays used in this study (see Dorman et al., in press).

The frequency upshifts for shorter insertions are less surprising than the outcome that longer insertions, for which there were still mismatches between filter CF and SG place, did not sound upshifted. Perhaps coming close in filter CF to the SG place frequency, when combined with strong top-down or prediction machine processing (Seth, 2019), is sufficient for this dimension—as it is for speech understanding (e.g., Dorman et al., 1997). On the other hand, it may be that the difference between filter CF and SG place frequency is not the appropriate metric to correlate with perceived spectral upshift.

This account assumes that the place pitch corresponding to an SG location is the relevant dimension for this discussion. However, the pitch percept for a given electrode (a) depends on stimulation rate (Eddington et al., 1978), (b) can change in both an apical and basal direction over time (e.g., Tan et al., 2017; Vermeire et al., 2015), and (c) is commonly lower that the SG frequency after some experience with electrical stimulation (e.g., Boëx et al., 2006; Reiss et al., 2007). Perceived pitch was not evaluated in this project, and it is not known if data of this kind would account for outcomes in a way that SG frequency did not.

Changes in F0 were not clearly related to insertion angle. One patient (#14), with the shortest insertion (394° or 680 Hz SG place frequency), appeared to play off a spectral upshift with a decrease in F0 to create a sound quality close to that of his CI. Similar outcomes, for patients with shorter insertions, have been described in Dorman, Natale, Baxter, et al. (2019) and Dorman, Natale, Zeitler, et al. (2019). This outcome appears to be cue trade-off because a decrease in F0 is unlikely to be the result of a shorter insertion per se.

A Complete Match

One patient indicated that the signal constructed for the NH ear was a complete match to the sound of her CI

and that signal was minimally changed relative to a clean signal. Because this outcome was so unlikely, we have followed the patient for 4 years. On her first visit in 2015, at age 15 years and 6 months following hookup, she scored 74% on AzBio sentences (Spahr et al., 2012) in quiet using the FS4p strategy and achieved a localization score of 15° of error (95th percentile of normal = 11°). Her match was constructed with minimal smear—a slight roll-off of high frequencies. She indicated that the match was a 10. Two years later (2017), she scored 82% correct on the AzBio sentences in quiet. Her match was constructed with a little more smear than previously and a −10 Hz shift in F0. Again, the she indicated the match was a 10.

The next year (2018) her AzBio score in quiet was 99% correct and the score at +5 dB signal-to-noise ratio was 57% correct. She scored 79% correct on consonants and 75% correct on vowels. Gifford et al. (2018) reported, for a large sample, a mean score for the AzBio sentences in quiet of 63% correct and 30% correct for sentences at +5 dB signal-to-noise ratio. Thus, performance by this patient was far above average. This level of performance may be related to her very high score, 88% correct, on a test of spectral modulation detection. Gifford et al. reported a mean score of 61% correct for approximately 500 CI patients on this test. It is relevant to note that her electrode insertion angle was 581° with a SG frequency of 300 Hz. This was a deep insertion but far from the deepest in our sample. The match was to a completely clean signal.

In 2019, she scored 83% correct on the AzBio in quiet, 46% correct in noise, 73% for consonants, and 95% for vowels. Her match was to a signal low passed at 3 kHz. Her match score was 10. Thus, over four visits the matches were between a slightly muffled signal and a clean signal. Finally, we note that the patient is a highly skilled musician and uses input to her CI, in addition to her NH ear, to tune her violin to other instruments in a quartet (see Landsberger et al., 2019, for a report that music sounds better for SSD-CI patients when the CI ear is included). While these outcomes seem very unlikely, we abide by the data.

What Is Missing From the Approximations?

After a patient gave a match score, she/he was asked "What is missing from the match?" The responses are shown in Table 3.

Two themes are apparent in the responses. One is that the approximations were slightly clearer than the CIs. The second is that the approximations needed to be more harsh, sharper, or harder. We return to this later.

During the many hours of testing on this and other projects, patients volunteered that the approximations lacked a 'robotic' or 'mechanical' sound quality. SSD-CI patients' percept of a robotic sound quality has been difficult to approximate by stimulation to the NH ear.A signal with a flat F0 contour sounds robotic to most NH listeners, but no patient has wanted F0 completely flattened to make a match. Signals created by a sine vocoder also sound robotic to NH listeners, but SSD-CI patients indicate that CI signals do not sound like the output of a sine vocoder (Dorman et al., 2017; Peters et al., 2018).

Table 3. Responses to Question "What Is Missing From CI Simulation?"

Subject	Match rating	What is missing?
1	8.5	NA
2	9.5	The match is missing the extra hum that comes after the end of words in the CI
3	9.0	Difficult to describe
4	9.0	Needs to be harder or more "punch"
5	10.0	Nothing
6	9.0	CI less clear
7	9.0	NA
8	9.5	NA
9	9.0	CI is less clear
10	8.0	CI more harsh, talking with mouth tightly closed
11	7.5	Missing distortion at the beginning and end of words.
12	7.0	CI is sharper
13	9.5	NA
14	8.5	Cannot quite describe difference

Note. NA = question not asked; CI = cochlear implant.

A musically trained listener volunteered that CI signals sound robotic because they lack the nuances of speech. From this point of view, we could improve our matches, that is, add a robotic sound quality, if we decimated the signals in the frequency domain. However, it is not clear which signal processing operations might accomplish this without sounding like a vocoder. Moreover, there may be perceptual consequences of electrical stimulation that are unique to electrical stimulation, for example, the perceptual consequences of the simultaneous, highly phase-locked firing of neural elements over a wide area of the cochlea. Indeed, this is a reasonable first guess for the mechanism underlying the percept of "hard."

The sound files contained in Supplemental Materials 2, 3 and 5 document that sound quality differs significantly among SSD-CI listeners. In addition, SSD-CI patients have reported vastly different perceptual experiences when listening to signals in their CI ear. At one end of the continuum, patients remark on the similarity of the experience of listening with their CI to listening with their NH ear. At the other end of the continuum, patients remark on how different the two perceptual

processes seem. For these patients, perception seems "immediate" for input to the NH ear. However, with input to the CI ear, perception seems slow and clearly constructive. For example, initial words in a sentence may not be understood and then are reconstructed when more information is available. This seemingly slow, reconstructive process appears to add another, not normal, dimension to sound quality for some SSD-CI listeners.

Do Conventional CI Patients Experience the Same Voice Quality as SSD-CI Patients?

The answer to this question depends on the answer to the question of how a normal cortical representation of speech (from the ear with normal auditory thresholds) influences the perception of voice for SSD-CI patients. One possibility is that the normal representation of speech emphasizes the abnormalities in signal representation engendered by electrical stimulation. On this view, conventional patients would experience a different and better voice quality than SSD-CI patients. On the other hand, a normal cortical representation of speech may serve as a reference from which the electrical representation can be normalized (e.g., Johnson, 2005). On this view, conventional patients, who lack a temporally immediate reference, would experience a poorer voice quality than SSD-CI patients. Finally, it is possible, but unlikely, that the normal representation of speech from the ear opposite the implant for SSD-CI patients has no effect on their perception of CI voice. On this view, our results would be generalizable to the conventional unilateral CI patient. In the absence of convincing evidence for any one of the three possibilities, this question remains open.

Limitations of Study

Our approach to matching the sound of a CI has been a "black box" approach because we do not know enough about the electrically elicited representation of speech at the periphery and in the cortex to generate an appropriate model. With a model, predictions about contributions to CI sound quality could be generated and then tested. We look forward to such a model.

The sound of a CI can be approximated by manipulating the characteristics of multiple acoustic dimensions. It is reasonable to suppose that patients could attend primarily to one dimension during one test session and to another dimension in another test session. If so, it will be difficult to replicate results.

We have not found an automated search algorithm for our dimensional space. For that reason, we used a single, very experienced, experimenter to manipulate the values of parameters for each of the stimulus dimensions. The matching scores of 9.0 and higher suggest that this was an effective way to derive a match. That said, the training that allowed the experimenter to arrive efficiently at a match, by definition, prevented a neutral, or unbiased, approach to the match. When asking an open-ended question, "What do I need to do to make the signal in your NH ear sound like your CI?" the value of an experienced experimenter likely exceeds the problems associated with experimenter bias. However, the use of a single, trained experimenter raises the issue of replicability.

Patients no doubt have different internal standards for saying a match is, for example, a 9 of 10. One patient's match score of 7 could be equivalent to another patient's score of 9. Even cultural differences could play a role in the value assigned to a match. As noted by one reviewer, a listener with experience in an educational system where high scores are not expected may volunteer lower scores than a listener with educational experience in the United States. That said, all but one of the patients in the Polish sample described earlier volunteered scores of 9.0 and 9.5.

Finally, the methods used in the research described here were far from psychophysically rigorous. For that reason, the outcomes should be viewed as a qualitative rather than quantitative description of sound quality.

Conclusions

It is possible to create a good approximation to the sound of a CI by altering a clean speech signal using a small number of operations, for example, low-pass filtering, band-pass filtering, spectral peak broadening, and F0 contour flattening. In this study, the approximations were, most generally, slightly more clear than the CI. The approximations varied greatly in sound quality suggesting that there large differences in CI sound quality among SSD-CI patients. In the best case, CI sound quality appears to be only slightly muffled.

Appendix: Descriptions of Sound Quality (Modified From https://www.complyfoam.com/the-newbies-guide-to-describing-sound/)

Aggressive
Airy
Bassy—emphasized bass
Blanketed
Bloated
Blurred
Boomy
Boxy
Breathy

Bright
Chipmunk-like
Clear
Closed
Congested
Colored
Cool
Computer-like
Crisp
Dark
Darth Vader-like
Delicate
Detailed
Distorted
Dry
Dull
Dynamic
Edgy
Far-away
Full
Grainy
Grip
Grungy
Hard
Harsh
High-pitched
Hollow
Laid-back
Lush
Mellow
Metallic
Mickey Mouse-like
Muddy
Muffled
Munchikin-like
Nasal
Open
Piercing
Punchy
Rich
Reverberant
Round
Shrill
Sibilant aka "Essy"
Sizzly
Smeared
Smooth
Spacious
Steely
Telephone like
Thick
Thin
Tinny
Treble-y
Veiled
Warm
Wet

Acknowledgments

Dan Freed of Advanced Bionics wrote the code for the metallic filter. This work could not have been accomplished without the assistance of Dr. John Karis and the neuroradiology staff at Barrow Neurological Institute in Phoenix, Arizona.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: M.F.D. is a consultant for Advanced Bionics and MED-EL Corporations. D.M.Z. is a consultant for Advanced Bionics and MED-EL Corporation. The other authors declare no conflict of interest.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was supported by grants from MED-EL Corporation to M.F.D., from Advanced Bionics to D.M.Z., and from National Institutes of Health R01DC014037, R01DC008408, R01DC014462, R01DC15798, R21DC16153, R01DC017683 to J.H.N.

ORCID iDs

Michael F. Dorman https://orcid.org/0000-0001-9560-3231
Daniel M. Zeitler https://orcid.org/0000-0001-5506-5483
Jeroen P. M. Peters https://orcid.org/0000-0003-3189-9103

Supplemental material

Supplemental material for this article is available online.

References

Arndt, S., Aschendorff, A., Laszig R., Beck, R., Schild, C., Kroeger, S., Ihorst, G., & Wesarg, T. (2011). Comparison of pseudo-binaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otology and Neurotology*, *32*(1), 39–47. https://doi.org/10.1097/MAO.0b013e3181fcf271

Baer, T., & Moore, B. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of the Acoustical Society of America*, *94*(3), 1229–1241. https://doi.org/10.1121/1.408176

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A, Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., . . . Lazard, D. S. (2013). Factors affecting auditory performance of post-linguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology*, *18*(1), 36–47. https://doi.org/10.1159/000343189

Boëx, C., Baud, L., Cosendai, G., Sigrist, A., Kós, M-I., & Pelizzone, M. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing, *Journal of the Association for Research in Otolaryngology*, *7*(2), 110–124. https://doi.org/10.1007/s10162-005-0027-2

Dorman, M. F., Loizou, P., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *Journal of the Acoustical Society of America*, *102*(5), 2993–2996. https://doi.org/10.1121/1.420354

Dorman, M. F., Natale, S., Baxter, L., Zeitler, D., Carlson, M., & Noble, L. (2019a). Cochlear place of stimulation is one determinant of cochlear implant sound quality. *Audiology and Neurotology*, *24*(5), 264–269. https://doi.org/10.1159/000503217

Dorman, M., Natale, S., Butts, A., Zeitler, D., & Carlson, M. (2017). The sound quality of cochlear implants: Studies with single-sided deaf listeners. *Otology and Neurotology*, *38(*8*)*, e268–e273. https://doi.org/10.1097/MAO.0000000000001449

Dorman, M., Natale, S., Zeitler, D., & Noble, J. (2019b). Looking for Mickey Mouse but finding a Munchkin: The perceptual effects of frequency upshifts for single-sided-deaf, cochlear implant patients. *Journal of Speech, Language and Hearing Research*, *15*, 1–7. https://doi.org/10.1044/2019_JSLHR-H-18-0389

Dorman, M., Spahr, A., Gifford, R., Loiselle, L., McKarns, S., Holden, T., Skinner, M., & Finley, C. (2007). An electric frequency-to-place map for a cochlear implant patient with hearing in the non-implanted ear. *Journal of the Association for Research in Otolaryngology*, *8*(2), 234–240. https://doi.org/10.1007/s10162-007-0071-1

Dorman, M., Zeitler, D., Cook, S., Loiselle, L., Yost, W., Wanna, G., & Gifford, R. (2015). Interaural level difference cues (ILDs) determine sound source localization by single-sided deaf patients fit with a cochlear implant. *Audiology and Neurotology*, *20*(3), 183–188. https://doi.org/10.1159/000375394

Eddington, D., Dobelle, W., Brackmann, D., Mladejovsky, M., & Parkin, J. (1978). Auditory prosthesis research with multiple channel intracochlear stimulation in man. *Annals of Otology, Rhinology and Laryngology*, *87*(6 Part 2), 1–39.

Gifford, R., Noble, J., Camarata, S., Sunderhaus, L., Dwyer, R., Dawant, B., Dietrick, M., & Labadie, R. (2018). The relationship between spectral modulation detection and speech recognition: Adult versus pediatric cochlear implant recipients. *Trends in Hearing*, *22*, 1–14. https://doi.org/10.1177/2331216518771176.

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni and R. Remez (Eds.), *The handbook of speech perception*. (pp. 363–369). Blackwell.

Karoui, C., James, C., Barone, P., Bakhos, D., Marx, M., & Macherey, O. (2019). Searching for the sound of a cochlear implant: Evaluation of different vocoder parameters by cochlear implant users with single-sided deafness. *Trends in Hearing*, *23*, 1–15. https://doi.org/10.1177/2331216519866029

Kawahara, H., Estill, J., & Fujimura, O. (2001, September 13–15). Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT. In 2*nd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA), Florence, Italy* (pp. 59–64). ICSA.

Landsberger, D., Svrakic, M., Roland, 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

Landsberger, D., Vermeire, K., Stupak, N., Lavender, E., Neukam, J, Van de Heyning, P., & Svirsky, M. (2019). Music is more enjoyable with two ears, even if one of them receives a degraded signal provided by a cochlear implant. *Ear and Hearing*. Advance online publication. https://doi.org/10.1097/AUD.0000000000000771

Loizou, P. (1998). Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants. *IEEE Signal Processing Magazine*, *15*(5), 101–130.

Loizou, P. (2006). Speech processing in vocoder-centric cochlear implants. In A. Møller (Ed.), Cochlear and brainstem implants (Vol. 64, pp. 109–143). Karger. https://doi.org/10.1159/000094648

Manrique, M., Calavia, D., Manrique-Huarte, R., Zulueta-Santos, C., Martin, M., & Huarte, A. (2019). Prelingual deaf children treated with cochlear implant: Monitoring performance with percentiles. *Otology and Neurotology*, *40*(5S Suppl 1), S2–S9. https://doi.org/10.1097/MAO.0000000000002206

Noble, J., Labadie, R., Gifford R., & Dawant, B. (2013). Image-guidance enables new methods for customizing cochlear implant stimulation strategies. *IEEE Transactions for Neural Systems and Rehabilitation Engineering*, *21*(5), 820–829. https://doi.org/10.1109/TNSRE.2013.2253333

Peters, J., Wendric, A., van Eijl, R., Rheberge, K., Versnel, H., & Grolman, W. (2018). The sound of a cochlear implant investigated in patients with single sided deafness and a cochlear implant. *Otology and Neurotology*, *39*(6), 707–714. https://doi.org/10.1097/MAO.0000000000001821

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

Riss, D., Hamzavi, J., Blineder, M., Honeder, C., Ehrenreich, I., Kaider, A., Baumgartner, W-D., Gstoettner, W., & Arnoldner, C. (2014). FS4, FS4-p, and FSP: A 4-month crossover study of 3 fine structure sound-coding strategies. *Ear and Hearing*, *35*(6), e272–e281. https://doi.org/10.1097/AUD.0000000000000063

Seth, A. (2019). The neuroscience of reality. *Scientific American*, September, 40–47. https://doi.org/10.1038/scientificamerican0919-40

Shannon, R., Zeng, F-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Sharma, A., Glick, H., Campbell, J., Torres, J., Dorman, M., & Zeitler, D. (2016). Cortical plasticity and reorganization in pediatric single-sided deafness pre- and post-cochlear

implantation. A case study. *Otology Neurotology*, *37*(2), e26–e34. https://doi.org/10.1097/MAO.0000000000000904

Spahr, A., Dorman, M., Litvak, L., van Wie, S., Gifford, R., Loiselle, L., Oakes, T., & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing*, *33*(1), 112–117. https://doi.org/10.1097/AUD.0b013e31822c2549

Svirsky, M., Ding, N., Sagi, E., Tan, C. T., Fitzgerald, M., Glassman, E. K., Seward, K., & Neuman, A. C. (2013). Validation of acoustic models of auditory neural prostheses. In IEEE International Conference on Acoustics and Speech Signal Processing (pp. 8629–8633). IEEE. https://doi.org/10.1109/ICASSP.2013.6639350

Tan, C.-T., Martin, B., & Svirsky, M. (2017). Pitch matching between electrical stimulation of a cochlear implant and acoustic stimuli presented to a contralateral ear with residual hearing. *Journal of the American Academy of Audiology*, *28*(3), 187–199. https://doi.org/10.3766/jaaa.15063

Tobey, E., Thal, D., Niparko, J., Eisenberg, L., Quittner, A., Wang, N. Y., & the CDaCI Investigative Team. (2013). Influence of implantation age on school-age language performance in pediatric cochlear implant users. *International Journal of Audiology*, *52*(4), 219–229. https://doi.org/10.3109/14992027.2012.759666

Van de Heyning, P., Vermeire, K., Diebl, M., Nopp, P., Anderson, I., & De Ridder, D. (2008). Incapacitating unilateral tinnitus in single-sided deafness treated by cochlear implantation. *Annals of Otology, Rhinology and Laryngology*, *117*(9), 645–652. https://doi.org/10.1177/000348940811700903

Vermeire, K., Landsberger, D., Van de Heyning, P., Voormolen, M., Kleine-Punte, A., Schatzer, R., & Zierhofer, C. (2015). Frequency-place map for electrical stimulation in cochlear implants: Change over time. *Hearing Research*, *326*, 8–14. https://doi.org/10.1016/j.heares.2015.03.011

Wilson, B., Dorman, M., Gifford, R., & McAlpine, D. (2016). Cochlear implant design considerations. In N. Young & K. I. Kirk (Eds.), *Pediatric cochlear implantation: Learning and the brain* (pp. 3–23). Springer.

Wilson, B., Finley, C., Lawson, D., & Zerbi, M. (1997). Temporal representations with cochlear implants. *American Journal of Otology*, *18*(Suppl. 6), S30–S34.

Zeitler, D., & Dorman, M. (2019). Cochlear implantation for single-sided seafness: A new treatment paradigm. *Journal of Neurological Surgery Part B*, *80*(2), 178–186. https://doi.org/10.1055/s-0038-1677482

Zeitler, D., Dorman, M., Cook, S., Loiselle, L., Yost, W., & Gifford, R. (2015). Sound source localization and speech understanding in complex listening environments by single-sided deaf listeners after cochlear implantation. *Otology and Neurotology*, *36*(9), 1467–1471. https://doi.org/10.1097/MAO.0000000000000841

Zeng, F-G., Tang, Q., & Lu, T. (2014). Abnormal pitch perception produced by cochlear implant stimulation. *PLoS One*, *9*(2), e88662. https://doi.org/10.1371/journal.pone.0088662