Musical Sound Quality Impairments in Cochlear Implant (CI) Users as a Function of Limited High-Frequency Perception

Alexis T. Roy, MSc[1], Patpong Jiradejvong, MSc[1], Courtney Carver, AuD[1], and Charles J. Limb, MD[1,2]

Abstract

The purpose of this study was to (a) apply the musical sound quality assessment method, Cochlear Implant-MUltiple Stimulus with Hidden Reference and Anchor (CI-MUSHRA), to quantify musical sound quality deficits in CI (cochlear implant) users with respect to high-frequency loss, and (b) assess possible correlations between CI-MUSHRA performance and self-reported musical sound quality, as assessed by more traditional rating scales. Five versions of real-world musical stimuli were created: 8-,4-, and 2-kHz low-pass-filtered (LPF) versions with increasing high-frequency removal, a composite stimulus containing a 1-kHz LPF-filtered version and white noise ("anchor"), and an unaltered version ("hidden reference"). Using the CI-MUSHRA methodology, these versions were simultaneously presented to participants in addition to a labeled reference. Participants listened to all versions and provided ratings based on a 100-point scale that reflected perceived sound quality difference among the versions. A total of 25 musical stimuli were tested. As comparison measures, participants completed four Visual Analogue Scales (VAS) to assess musical sound quality. Overall, compared to normal hearing (NH) listeners, CI users demonstrated an impaired ability to discriminate between unaltered and altered musical stimuli with variable amounts of high-frequency information removed. Performance using CI-MUSHRA to evaluate this parameter did not correlate to measurements of musical sound quality, as assessed by VAS. This study identified high-frequency loss as one acoustic parameter contributing to overall CI-mediated musical sound quality limitations. CI-MUSHRA provided a quantitative assessment of musical sound quality. This method offers the potential to quantify CI impairments of many different acoustic parameters related to musical sound quality in the future.

Keywords

cochlear implants, sound quality, music perception

Introduction

Although many cochlear implant (CI) users exhibit remarkable success with respect to speech comprehension, music perception remains essentially poor for most CI users. In addition to clear deficits in pitch discrimination, and melody and timbre identification (Limb & Rubinstein, 2012; McDermott, 2004), CI users also report musical sound quality impairments following implantation (Lassaletta et al., 2008a). The quality of music is frequently perceived as unpleasant, mechanical (unnatural), lacking fullness, and/or rough, for example (Looi, Winter, Anderson, & Sucher, 2011; Mirza, Douglas, Lindsey, Hildreth, & Hawthorne, 2003). In the case of music, sound quality (i.e., how good the music sounds) may indeed be the single most significant factor responsible for a listener's enjoyment (Lassaletta et al., 2008b). Despite its importance for music perception, however, few studies have rigorously addressed musical sound quality deficits outside of questionnaire-based assessments. Instead, the focus has been placed on developing tests that

assess the accuracy of CI users to perceive music in its deconstructed form, such as melody recognition, pitch discrimination, and rhythm identification assessments. These perceptual accuracy-based tests have become increasingly more standardized with direct aims of clinical use (Brockmeier et al., 2011; Kang et al., 2009). In comparison, few standardized methods exist to quantify perception of musical sound quality. This is a concern because no clear relationship exists between perceptual accuracy-based performance and the perceived sound quality of music (Gfeller et al., 2008; Looi et al., 2011). For example, the ability to identify a melody does not necessitate that the perceived

[1]Johns Hopkins Hospital, Baltimore, MD, USA
[2]Peabody Conservatory of Music, Baltimore, MD, USA

Corresponding Author:
Charles J. Limb, Department of Otolaryngology- Head and Neck Surgery, Johns Hopkins Outpatient Center, 6th Floor, 601 North Caroline St, Baltimore MD 21288, USA
Email: climb@jhmi.edu

sound quality of that melody is satisfactory. Considering this distinction, more rigorous sound quality assessments are warranted.

Musical sound quality has most commonly been evaluated using questionnaire-based assessments and subjective ratings scales, such as Visual Analogue Scales (VAS; Gfeller, Christ, Knutson, Witt, & Mehr, 2003; Lassaletta et al., 2008a; Looi et al., 2011; Migirov, Kronenberg, & Henkin, 2009), where CI users are asked to rate the perceived quality of a sound using either a scale delineated by descriptive adjectives (e.g., *natural/mechanical, pleasant/unpleasant, poor/excellent*) or a scale related to preference (e.g., *like-dislike*). While such scales may be highly intuitive for some CI users, there are several limitations to these rating scale methods when trying to derive a meaningful understanding of the factors that contribute to impairments in musical sound quality. For one, a participant's appraisal ratings can be biased by a variety of factors unrelated to the sound quality perceived through the CI device. Music psychology studies reveal that variables such as level of musical training, demographic differences (e.g., age, gender), previous familiarity with a musical style, and personality traits (e.g., introvert vs. extrovert) can all influence ratings (Ginocchio, 2009; Kemp, 1996; LeBlanc, 1982, 1991; Nater, Abbruzzese, Kreb, & Elhert, 2006). As a result, it becomes difficult to tease apart the influence of these outside variables from the true effect of the sound quality deficits inherent to CI-mediated listening. In addition, large variability is to be expected based on how participants interpret the adjectives that set absolute values of the scales (Gfeller et al., 2000). To compound matters, many of the descriptors used (e.g., rough, dull) are unfamiliar to a population of nonmusicians, leaving participants with limited, or at times arbitrary, reference points to determine where a musical stimulus will fall on the scale in relation to the provided adjectives.

Due to the constraints described, in a prior study we modified a sound quality assessment method used in the audio industry, called MUltiple Stimulus with Hidden Reference and Anchor (MUSHRA), for CI use, which we renamed as CI-MUSHRA (ITU-R.Recommendation.BS.1534-1, 2003—International Telecommunication Union, 2003; see also Roy, Jiradejvong, Carver, & Limb, 2012). In the CI-MUSHRA method, a set of sound quality versions of a musical stimulus are created in which a specific acoustic feature of music (hypothesized to contribute to CI-mediated sound quality impairments) is degraded to various degrees among versions. The degree of degradation systematically varies, such that a range of sound quality versions exists from an unaltered version (i.e., "reference") to a highly degraded version (i.e., "anchor"). These versions in addition to a labeled reference (unaltered version for comparison measures) are simultaneously provided to the participant in a computer interface. The participant is required to listen to each version and provide numeric ratings (between 0 and 100) that reflect perceived sound quality differences among the version and labeled reference. The sound quality impairment due to the acoustic parameter under test can be quantified based on the amount of degradation needed to elicit a detrimental impact in quality ratings.

The CI-MUSHRA method offers several improvements over more traditional sound quality assessments. One of the primary advantages of this method is that intrinsic participant factors that may influence their ratings (such as musical genre preference or previous familiarity) are reduced in impact, since ratings reflect relative sound quality differences perceived between an unaltered musical stimulus and sound quality versions of that same stimulus. Furthermore, participants provide ratings based on their inherent understanding of sound quality using a scale from "0" (*very poor*) to "100" (*excellent*). Participants are not required to relate sound quality to unfamiliar adjectives in order to provide a rating. CI-MUSHRA also affords researchers the opportunity to quantify specific features of music that contribute to sound quality impairments based on the acoustic parameter chosen to degrade the stimuli.

In a previous study, we used CI-MUSHRA to conclude that bass (i.e., low) frequency impairments contribute to overall CI-mediated musical sound quality deficits: CI users demonstrated an impaired ability to detect sound quality difference between unaltered musical stimuli and stimuli missing up to 400 Hz of bass information. Surprisingly, on 7.3% of the CI-MUSHRA trials, CI users rated the anchor—1- to 1.2-kHz band-pass-filtered version—as having identical sound quality to the original, unfiltered musical stimulus (Roy, Jiradejvong, Carver, & Limb, 2012). These results indicated that CI users have difficulty perceiving sound quality impairments not only as a function of bass frequency loss but also suggested the possibility that high-frequency loss (above at least 1.2 kHz) was also a contributing factor that deserved further examination using the CI-MUSHRA approach.

Since the CI device has been optimized for speech understanding, the incoming acoustic wave is band-pass-filtered at a frequency range corresponding to human voices. As a result, frequencies above 8 kHz are not readily transmitted by the device. In contrast to speech, however, music readily contains instruments that transmit frequencies well above 8 kHz (Snow, 1931). The presence of these higher frequencies contributes to overall satisfactory musical sound quality for normal hearing (NH) listeners, who can detect frequencies up to 20 kHz (Roy et al., 2012). We anticipate the loss of these high frequencies has a detrimental impact on musical sound quality for CI users. To study this effect, real-world music stimuli were low-pass-filtered to remove increasing amounts of high-frequency information. The CI-MUSHRA method was then used to evaluate the ability of CI users to perceive alterations in sound quality as a function of high-frequency content. We hypothesized that CI users would demonstrate an impaired ability to make sound quality discriminations among stimuli missing variable amounts of high-frequency information.

Another key consideration of this study was to assess possible correlations between performances in CI-MUSHRA and self-reported musical sound quality, as evaluated through more traditional rating scales. Prior to the study, CI users were asked to complete a questionnaire containing four VAS delineated by the descriptions, "doesn't sound like music/sounds like music," "poor sound quality/excellent sound quality," "mechanical/natural," and "hard to follow/easy to follow." Considering the large subjectivity inherent in these quality ratings scales (independent of sound quality actually transmitted by the CI device), we hypothesized that weak correlations (if any) would be found between the more psychophysical sound quality measures obtained by CI-MUSHRA and self-reported music sound quality ratings obtained by VAS methods. If a lack of correlation between methods exists, it provides evidence that CI-MUSHRA offers a useful measure of sound quality not currently captured by subjective rating scales alone.

Studies with NH listeners suggest that previous exposure to a song can improve quality ratings (LeBlanc, 1982; Schuessler, 1948). Unlike NH listeners, however, CI users do not demonstrate a familiarity effect on quality ratings—a finding possibly attributable to the impaired ability of CI users to recognize songs, such that accurate assessment of familiarity is difficult (Gfeller et al., 2003; Looi et al., 2011). As stated previously, the CI-MUSHRA method notably differs from more traditional rating scales in that it requires participants to rate perceived differences among sound quality versions of the same musical stimulus. In other words, each rating of a musical stimulus version is relative to its unaltered, best sound quality version. Therefore, biases resulting from certain variables intrinsic to the participant (e.g., familiarity toward a stimulus) and independent of perceived sound quality should be lessened, as these biases should be equally distributed among all sound quality versions of a musical stimulus (e.g., participants would be equally familiar with all versions within a trial). To examine this issue further, we also assessed the impact of previous familiarity on sound quality ratings for both participant groups.

Method

Study Participants

Twelve postlingually deafened CI users (mean age = 49.6 ± 10.9 years) and 12 NH listeners (mean age = 43.8 ± 11.6 years) participated in this study. CI users had a mean length of 4.4 ± 4.1 years of implant experience and utilized a variety of devices and processing strategies, as shown in Table 1. No participant had musical training beyond an amateur level. All CI participants were clinically diagnosed as profoundly deaf in both ears prior to implantation. All participants were recruited at the Johns Hopkins Listening Center in Baltimore, Maryland. All experimental protocols were formally approved by the Institutional Review Board at Johns Hopkins Hospital. Informed consent was obtained for all participants.

Musical Background Questionnaire

NH listeners and CI users were asked to indicate their highest level of musical training and average number of hours per week spent listening to music. All CI users completed a questionnaire to assess their musical enjoyment and subjective appraisal of musical sound quality. To assess musical enjoyment after receiving the implant, CI users answered the question, "I would describe myself as a person who enjoys music a lot" using a Likert-type scale between 1 (*completely disagree*) and 5 (*completely agree*). CI users were asked to reflect back on their musical listening experiences and rate the quality of music via four 100-mm VAS set by the following dichotomous descriptive adjectives "doesn't sound like music/sounds like music," "very poor sound quality/excellent sound quality," "mechanical/natural," and "hard to follow/easy to follow." Questions for subjective appraisal of music quality were inspired by Gfeller et al. (2000) and Lassaletta et al. (2008a, 2008b). Consonant-Nucleus-Consonant with Phonemes (CNC-P) and Consonant-Nucleus-Consonant with Words (CNC-W) speech scores were collected from CI users' most recent audiologist visit (within at least 12 weeks prior to participation in the study).

Musical Stimuli

A total of 25 real-world musical songs were selected from the genres of classical, country, hip-hop, jazz, and rock. The pool of stimuli was chosen to provide a diverse representation of Western music. To assess possible effects of song familiarity on sound quality ratings, three popular songs (i.e., assumed to be well known to a general audience of nonmusicians) and two obscure songs were selected for each genre. For the country, hip-hop, and rock genres, a song was classified as popular if it was listed on either the "Billboard Top 100," "VH1's Greatest Rock/Hip-Hop/Songs of All times," or "Country Music Television 40 Greatest Songs" musical ranking sources. A song was classified as obscure if it did not appear on any ranking chart. For selection of popular and obscure songs within the classical and jazz genres, a faculty member of the Peabody Conservatory of Music (Baltimore, MD) was consulted. Song files were rendered as lossless versions from commercially available compact disc (CD) recordings with no initial modifications.

For each song, a 5-s excerpt was selected to serve as the reference stimulus. Four sound quality versions of each reference were created: a set of (8-, 4-, and 2-kHz) LPF cutoff versions containing increasing amounts of high-frequency loss, and a composite stimulus composed equally of a 1-kHz LPF cutoff of the reference and white noise (referred to as the "anchor"). Within each 5-s segment, the first 500 ms included a gradual increase in volume (i.e., fade-in effect)

Table 1. Cochlear Implant (CI) User Demographics

Participant	Sex	Age	Etiology	Profound deafness (years)	CI experience (years)	Device	Processor
CI1	F	58	Idiopathic	18	2.8	Med-El SonataTi100	Opus II
CI2	M	61	Meniere's	4	3.7	Med-El SonataTi100	Opus II
CI3[a]	F	47	Autoimmune	9	3	Med-El SonataTi100	Opus II
CI4[a]	M	35	Idiopathic	4	2.3	ABC HiRes 90K	Harmony
CI5	M	58	Meniere's	11	11	ABC Clarion	Harmony
CI6[a]	M	57	Meniere's	3	2.2	Med-El SonataTi100	Opus II
CI7[a]	F	56	Idiopathic	7	4.9	ABC HiRes 120	Harmony
CI8[a]	M	51	Idiopathic	15	14.3	ABC HiRes 120	Harmony
CI9	F	47	Autoimmune	1	1	Med-El SonataTi100	Opus II
CI10[a]	F	58	Autoimmune	15	2	ABC Clarion CII	Harmony
CI11	F	26	Meniere's	8	3.7	CC Nucleus 22	Freedom
CI12[a]	F	41	Idiopathic	4	1.5	ABC HiRes 90K	Harmony

Note: ABC = Advanced Bionics Corporation; CC = Cochlear Corporation; F = female; M = male.
[a]These participants were bilateral CI recipients; demographics correspond to implant side with that was tested in this study.

and the last 500 ms included a gradual decrease in volume (i.e., fade-out effect).

Frequency range selection for the LPF cutoffs were based on the average frequency bandwidth assignments for CI channels for each device type (Figure 1). Based on the electrode frequency maps, it was anticipated CI users would have some difficulty distinguishing sound quality as a function of high-frequency content for the unaltered version (reference), 8-kHz LPF cutoff version and possibly 4-kHz LPF cutoff version due to similar channel stimulation. In comparison, the anchor and 2-kHz LPF cutoff version should elicit significantly different sound quality ratings than the other stimulus versions.

Sound quality versions were created using Myfilter.vi run on LabVIEW (National Instruments, Austin, TX) and Audacity 1.3.7 Beta. All stimuli were normalized by root-mean-square power with equal-loudness contour adjustments using Adobe Audition 3.0 (San Jose, CA).

Test Procedure

Listening Environment

Participants sat in a soundproof booth and free-field stimuli were presented via a single calibrated loudspeaker (Sony SS-MB150H) positioned in front of the listener. The volume was preset to 70 decibels sound pressure level (SPL). Participants controlled stimuli presentation and provided answers using a computer touchscreen positioned directly in front of them.

For unilateral CI users, the nonimplanted ear was occluded with an ear plug to reduce the effect of any possible residual acoustic hearing. No hearing aids were used in either ear during the study. For uniformity, bilateral implant users selected the implant side for their first implant and occluded the contralateral ear with an ear plug. NH listeners utilized both ears during the study to help ensure ecologically valid testing

conditions (i.e., NH listeners are accustomed to binaural music listening conditions).

Familiarity Assessment

Prior to the sound quality rating portion of the study, we tested previous familiarity with the musical stimuli for each participant. Using a novel, computer-controlled interface designed in LabVIEW (National Instruments, Austin, TX) participants listened to each of the 25 reference versions one at a time in randomized order. Once the participant listened to the song segment, (s)he was instructed to answer the question, "Are you familiar with this song segment?," by clicking either "Yes" or "No" on the computer screen. Before a decision was recorded, participants had the opportunity to replay each segment as many times as they wished. Participants relied solely on auditory cues and memory to assess familiarity, as no title or artist information was provided. Songs within the rock, country, and hip-hop genres contained lyrics, which could presumably help facilitate listeners to recognize these segments.

CI-MUSHRA Training

After assessment of familiarity, participants completed a brief training session in which they were exposed to the nature and range of sound quality variations among test stimuli. Using an interactive computer-controlled system inspired by Vincent (2005), participants were presented with a subset of the musical stimuli, which included one representative labeled-reference stimulus from each genre and its corresponding sound quality versions (8-, 4-, 2-kHz LPF cutoff versions, and anchor) in random order. Participants were required to play each labeled-reference and its sound quality versions at least once by clicking the appropriate icons. They were instructed to listen closely for differences in sound quality among versions. Written instructions were
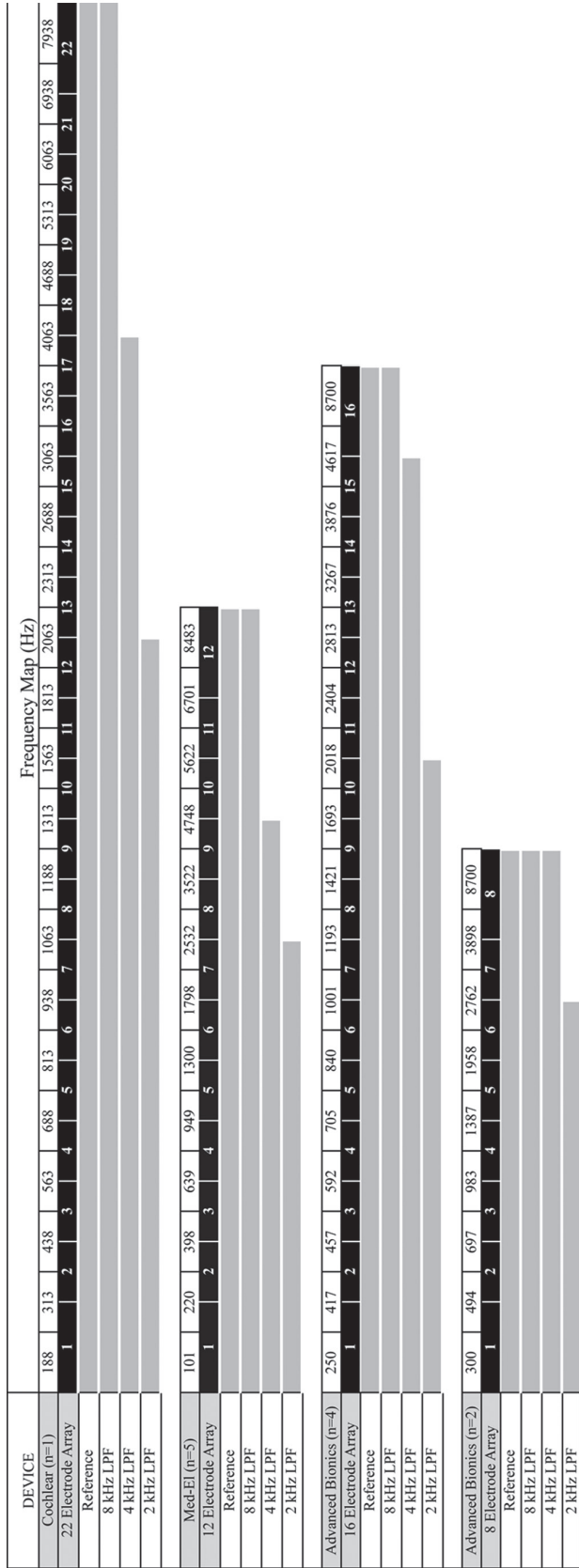
Figure 1. Average electrode frequency maps for cochlear implant (CI) users

Note: Within the speech processor, the incoming acoustic wave is band-pass-filtered into separate channels so that each electrode can later receive a specific frequency range of the original signal. The electrode map shown here indicates the average frequency range assigned to each electrode for CI participants of this study. Maps are separated by implant manufacture and number of physical electrodes within the array. The gray-scale bars represent anticipated electrode stimulation for each stimulus version. Electrically, there is little difference between the reference and 8-kHz LPF (low-pass-filtered) versions, and greater differences are anticipated for the 4- and 2-kHZ LPF versions.

freely provided. Each stimulus could be replayed without limit.

CI-MUSHRA Evaluation

Participants completed a total of 25 CI-MUSHRA evaluations, one for each song segment. Trials were randomized among participants. Within each CI-MUSHRA evaluation, participants were presented with a labeled reference (i.e., unaltered, best sound quality version) and five sound quality versions of the reference presented in random order, which included a "hidden reference" (i.e., identical to the labeled reference), 8-, 4-, and 2-kHz LPF cutoffs of the reference, and the anchor. Participants were instructed to listen to each stimulus at least once and provide a sound quality rating for each of the five versions using a 100-point scale. The rating scale was divided into five equal intervals, delineated by the adjectives *very poor* (0-20), *poor* (21-40), *fair* (41-60) *good* (61-80), and *excellent* (81-100). Written instructions were freely provided. See Figure 2 for a screen shot of the CI-MUSHRA evaluation interface used in this study.

More important, participants were instructed to provide ratings that reflect perceived sound quality differences among the labeled reference and all stimuli under test, as opposed to rating the absolute sound quality of any single test stimulus. It was required that at least one stimulus within the set be rated a score of 100 points, which indicated that this stimulus had identical sound quality to the labeled reference. In theory, the hidden reference should be assigned this score, since no high-frequency information was removed from this version. The evaluator can track the ability of participants to make fine discriminations among stimuli missing small amounts of high-frequency information by whether the hidden reference received a rating of 100.

Participants were not required to assign any stimulus in the set a score of 0 points, if she or he did not perceive that the sound quality of any of the stimuli were degraded enough to warrant this score. However, the purpose of the anchor (highly degraded sound quality version) was to help ensure utilization of the lower range of the rating scale, which in doing so would further increase rating resolution among LPF cutoff stimuli by distributing ratings along the entire 100-point scale. If participants perceived that two or more stimuli within a trial were identical in sound quality, they were allowed to assign these stimuli the same rating. In other words, ties in ratings were allowed between stimuli within a trial. Participants had the option to listen to any stimulus in any order without limit, which presumably would further facilitate a high degree of resolution among ratings.

The training and evaluation interface was configured and run using MATLAB Version 7.2 (The MathWorks, Inc., Natick, Massachusetts) and LabVIEW (National Instruments, Austin, TX). For a more extensive review of the CI-MUSHRA methodology, please refer to Roy et al. (2012). For more
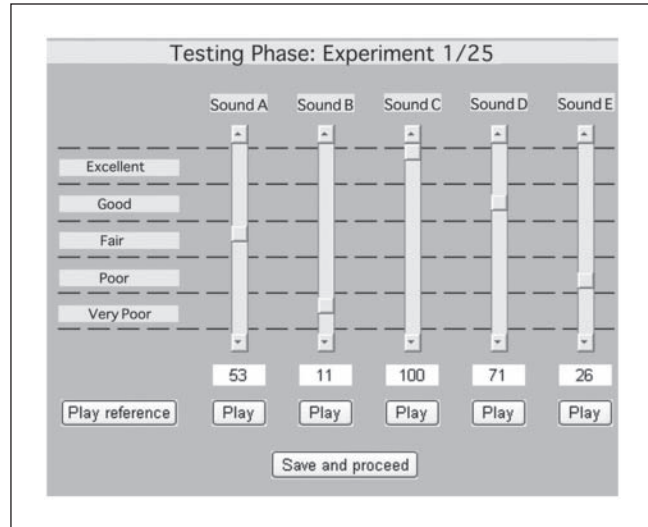
Figure 2. Screenshot of the CI-MUSHRA (Cochlear Implant-MUltiple Stimulus With Hidden Reference and Anchor) computer interface

Note: For each trial, participants are presented with the labeled reference (unaltered musical stimulus; marked "Play Reference") and five sound quality versions of the reference in random order (8-, 4-, 2-kHz LPF [low-pass-filtered] versions, the anchor, and hidden reference; marked as "Sounds A-E"). Participants can listen to each of these versions without limit by pressing "Play." To rate a stimulus, participants move the adjustable sliders between 0 and 100. Participants must rate at least one stimulus in the set a score of "100" (i.e., identical sound quality to the labeled reference). Once participants are satisfied with their ratings, they can press "Save and proceed" to move to the next trial. (Note that this figure differs from the figure in Roy, Jiradejvong, Carver, and Limb [2012] in that this interface has fewer sound quality versions per trial to allow for a shorter test paradigm.)

information on MUSHRA, please refer to ITU-R. Recommendation.BS.1534-1, 2003.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistic 19 (Somers, NY) with an alpha level of .05.

Results

There was no significant difference in age between NH listeners and CI users (two-tailed independent Student's *t* test, $t = 1.27$, $df = 22$, $p = .22$). No significant difference existed between groups for self-reported music listening habits (Mann–Whitney *U* test, $U = 71.5$, $p = .98$) or level musical training (Mann–Whitney *U* test, $U = 68$, $p = .81$).

Sound quality ratings provided by NH listeners and CI user for each stimulus version are displayed in Figure 3. A two-way ANOVA (analysis of variance) examined the effects of hearing classification (NH listener or CI users) and stimulus version (hidden reference, 8-, 4-, and 2-kHz LPF cutoff, and anchor) on sound quality ratings. A significant interaction between the effects of hearing classification
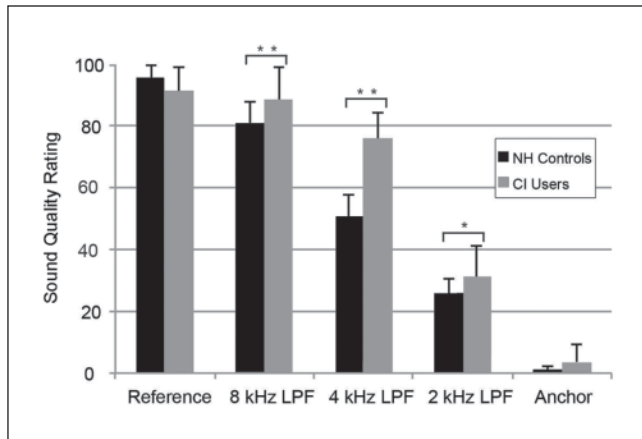
Figure 3. Overall mean sound quality ratings for each stimulus version by group

Note: Error bars indicated 1 standard deviation of the mean.

*$p = .06$ (indicates difference approached significance; two-way ANOVA [analysis of variance]).

**$p < .01$ (indicates difference between groups was significant; two-way ANOVA).

and stimulus version on sound quality ratings was found, $F(4, 110) = 14.3$, $p < .001$. To further examine this interaction, a simple effects test for hearing classification at each level of the stimulus version was conducted. NH listeners and CI users differed in their sound quality ratings for the 8-kHz LPF cutoff version ($p < .01$) and the 4-kHz LPF cutoff version ($p < .01$); differences in sound quality ratings between groups approached significance for the 2-kHz LPF cutoff versions ($p = .06$).

A simple-effects test for stimulus version at the level of NH listener and CI user was performed. At the level of NH listener, a pairwise comparison for stimulus versions revealed that all possible contrasts were significant ($p < .01$; Tukey's Honestly Significant Difference [HSD] pairwise comparison), providing evidence that NH listeners could distinguish sound quality difference among all stimulus versions. In comparison, overall, CI users did not assign significantly different sound quality ratings among the reference and 8-kHz LPF ($p = .93$; Tukey's HSD). All other contrasts for the CI user groups, as a whole, were significantly different ($p < .01$; Tukey's HSD). Taken together, compared to NH listeners, CI users demonstrated an impaired ability to make sound quality discriminations as a function of high-frequency content.

Due to large variability typically found in performance for CI users during music perception tasks, post hoc pairwise comparisons between stimulus versions were individually conducted for each CI user. This analysis revealed that CI users could be separated into two groups based on performance. As much as 42% of the CI users (5 out of 12) were unable to consistently make sound quality discrimination among the reference, 8- and 4-kHz LPF versions ($p > .05$; Tukey's HSD), providing an indication that this group of CI

users are missing at least 4 kHz of high-frequency information with respect to musical sound quality. All other contrasts for this group were significantly different ($p < .01$; Tukey's HSD). The remaining 58% of CI users (7 out of 12) were more adept at making sound quality discriminations as a function of high-frequency content and more closely resembled performance by NH listeners. For this group, all contrasts were significantly different ($p < .01$), except for sound quality ratings between the reference and 8-kHz LPF cutoff version ($p > .05$). There was no significant difference in age, CI device manufacture, hearing loss duration, CI experience (i.e., years of CI use), or musical experience between these two groups of CI users (Student's t test, $p > .05$).

Sound quality ratings were plotted as a function of LPF filter cutoff (kHz). The anchor was excluded from the data set, since it was parametrically different from the other stimuli (i.e., contained white noise in addition to the LPF). The reference was treated as a 20-kHz LPF cutoff for data analysis purposes (i.e., the typical LPF cutoff for NH). Logarithmic and linear regression models were tested for their fit to sound quality ratings for NH listeners and CI users. The coefficient of determination (i.e., R^2) indicated that a logarithmic regression model provided a better fit for the data than a linear regression model: 96% of variance in sound quality ratings was predicted from LPF cutoff using a logarithmic model, as compared to only 77% with a linear model for NH listeners. Comparatively, for CI users the data were better represented by a logarithmic regression fit ($R^2 = .75$) than linear regression fit ($R^2 = .42$) for sound quality ratings.

Figure 4 represents the best-fit line for NH listeners (Y = $31.1 \times \text{Ln}(x) - 207.3$) and CI users (Y = $24.6 \times \text{Ln}(x) - 142.0$). For this data set, the slope of the best-fit line quantifies the rate at which sound quality ratings increase with the addition of high-frequency information. Therefore, the more adept a participant is at making sound quality discriminations as a function of high-frequency content, the greater the slope of their line of best fit should be. The slope for NH listeners was significantly greater than the slope for CI users (two-tailed independent Student's t test, $t = 0.04$, $df = 22$, $p < .001$), providing further indication of CI users' difficulty with this task.

The slope of the best-fit line was individually calculated for each of the CI users to serve as a single measure of each participant's performance in this task. These slopes were then tested for correlation to CI users' speech scores and musical questionnaire answers. Average speech scores for Consonant-Nucleus-Consonant with words (CNC-W) and phonemes (CNC-P) were $68.2 \pm 18.3\%$ and $81.8 \pm 11.8\%$, respectively. There was no significant correlation between speech scores and performance in CI-MUSHRA (CNC-W: $r = .03$, $p = .93$; CNC-P: $r = .19$, $p = .56$).

On average, CI users reported 3.25 ± 1.5 points using the 5-point Likert-type scale to assess musical enjoyment. Self-reported musical enjoyment was not significantly correlated
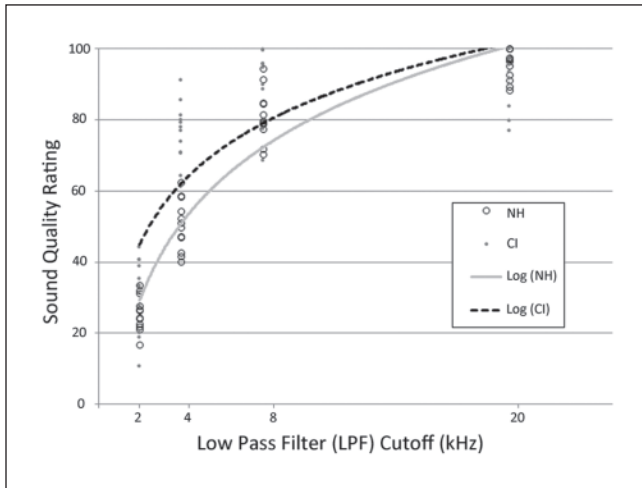
Figure 4. Logarithmic regression analysis of mean sound quality rating versus low-pass-filter (LPF) cutoff (kHz) by group
Note: NH = normal hearing; CI = cochlear implant. Data points correspond to average sound quality ratings for a stimulus version for one participant (best-fit line for NH listeners: $Y = 31.1 \times Ln(x) - 207.3$, $r^2 = .92$; best-fit line for CI users: $Y = 24.6 \times Ln(x) - 142.0$, $r^2 = .75$).

to performance in CI-MUSHRA ($r = -.09$, $p = .77$). For the subjective musical appraisal scales, CI users on average rated "doesn't sound like music/sounds like music" as 70.4 ± 26.7 points; "very poor sound quality/excellent sound quality" as 58.7 ± 29.5 points; "mechanical/natural" as 57.8 ± 26.0; and "hard to follow/easy to follow" as 57.7 ± 31.3 points. None of the subjective appraisal scores correlated significantly to CI-MUSHRA performance (sounds like music: $r = -.31$, $p = .32$; sound quality: $r = -.508$, $p = .080$; mechanical/natural: $r = -.518$, $p = .084$; easy/hard: $r = -.392$, $p = .21$).

On average, CI users reported 4.75 ± 5.3 years of formal musical training. CI-MUSHRA performance was strongly correlated to amount of previous musical training received ($r = .67$, $p < .05$). CI-MUSHRA performance showed no significant correlation to time spent listening to music ($r = .20$, $p = .54$), age ($r = -.25$, $p = .44$), years of CI use ($r = .24$, $p = .46$), length of profound deafness ($r = -.04$, $p = .90$), LPF cutoff of speech processer ($r = -.44$, $p = .15$), or average frequency range assignment per electrode ($r = -.14$, $p = .67$),

With respect to previous familiarity, NH listeners were on average familiar with 11.3 ± 4.8 of the 25 musical song segments, as compared to 6.9 ± 4.4 for CI users. This difference was not significant between groups (two-tailed independent Student's t test, $t = 1.65$; $df = 22$; $p = .12$). A two-way ANOVA (familiarity and stimulus versions factors) was conducted for NH listeners and CI users to assess whether previous familiarity with the musical stimuli affected sound quality ratings. Familiarity had an insignificant effect on quality ratings for both NH listeners, $F(1, 1490) = 2.3$, $p = .13$, and CI users, $F(1, 1490) = 0.16$, $p = .69$.

Discussion

Overall, CI users demonstrated a greater difficulty in detecting sound quality difference as a function of high-frequency content as compared to NH listeners. As much as 58% of the CI users tested were unable to detect sound quality differences among the unfiltered stimuli and stimuli missing above 8 kHz of high-frequency information. For the remaining CI users, the impairment extended to stimuli missing above 4 kHz of high-frequency information. In contrast, NH listeners reliably and consistently made sound quality discriminations among all LPF cutoff stimuli. Ratings by NH listener reveal the importance of high frequencies for musical sound quality perception and further highlight the detrimental effect that high-frequency impairments may have on CI-mediated sound quality perception.

One aim of this study was to compare how sound quality ratings obtained by CI-MUSHRA compared to ratings collect by more traditionally used VAS. Overall, a very weak correlation was found between CI-MUSHRA performances and self-reported subjective ratings. More specifically, CI-MUSHRA scores were insignificantly correlated to quality ratings collected by the following VAS: "sounds like music/doesn't sound like music," "excellent sound quality/very poor sound quality," "natural/mechanical," and "easy to follow/hard to follow." These results suggest that CI-MUSHRA performance is largely independent of subjectively reported sound quality (as quantified by the VAS used here) and provides evidence that CI-MUSHRA can offer a nonredundant measure of musical sound quality perception.

Discrepancies between an individual's performance in CI-MUSHRA and his or her subjectively reported ratings collected by VAS methods highlight the usefulness of CI-MUSHRA as a quantitative measure of musical sound quality. For example, participant CI5 demonstrated one of the poorest sound quality discrimination abilities of any CI user, but subjectively reported very high musical sound quality perception (89/100) and good musical enjoyment postimplantation (4/5). Using traditional questionnaire-based measures, it could be anticipated that participant CI5 has satisfactory sound quality perception. In the testing reported here, participant CI5 perceived a highly degraded musical stimulus (missing > 4 kHz of frequency information) as identical to an unaltered stimulus, indicating an extreme deterioration in sound quality. The high subjective ratings of CI5 most likely reflect the ability of some CI users to adapt to the degraded musical information as a new standard for musical sound quality. Nevertheless, sound quality perception is far from normal in most CI users. CI-MUSHRA can provide a sensitive measure of these sound quality limitations.

In this study, we found the following variables—years of CI use, length of profound deafness, speech scores, and speech processor type—were not associated with performance in CI-MUSHRA. These results are in agreement with

a previously reported study that assessed relationships between such variables and subjective music appraisal using questionnaire-based formats (Migirov et al., 2009). Interestingly, however, years of musical training (either before or after implantation) were positively correlated with performance in CI-MUSHRA. There is accumulating evidence to suggest that training can positively affect music perception abilities (Driscoll, Oleson, Jiang, & Gfeller, 2009; Galvin, Fu, & Shannon, 2009; Gfeller et al., 2000, 2002). For example, Gfeller et al. (2002) reported that the perceived quality of musical instruments significantly improved following a training program. It is possible that CI users with musical training are better at reconstructing the degraded musical stimuli presented by the device, such that they have improved musical sound quality perception.

As compared to more traditional rating scales, CI-MUSHRA is notably different in that it requires listeners to detect sound quality differences between altered versions of the same musical stimulus. As a result, we anticipate that variables unrelated to perceived sound quality that may bias ratings (e.g., previous familiarity, genre preferences) should be uniformly distributed among versions within a trial and thus have little effect on CI-MUSHRA ratings. We found here that the interparticipant variable of previous familiarity did not impact sound quality ratings obtained by the CI-MUSHRA methodology for both NH listeners and CI users. Similarly, other variables that may bias sound quality ratings (e.g., musical style preference) are less likely to have significant impact using the CI-MUSHRA approach, although further study is warranted.

As previously described, the anchor is a severely degraded sound quality version of the labeled reference. Its role is to help ensure participants utilize the lower range of the rating scale (ITU-R.Recommendation.BS.1534-1, 2003). In an earlier CI-MUSHRA study, a 1- to 1.2-kHz band-pass-filtered version of the original musical stimulus was not degraded enough in quality to consistently receive low ratings by CI users (Roy et al., 2012). In this present study, the anchor was a 1-kHz LPF of the reference in which white noise was added to ensure extreme degradation. CI users on average rated the anchor, 3.7 ± 5.4 points, which were lower and less variable than anchor ratings in the previous study (31.3 ± 36.0 points). These results suggest that the anchor used here better served its intended purpose: It was degraded enough in sound quality for CI users to detect that it was significantly worse than the other versions. However, the anchor was parametrically different than the other sound quality versions, which makes anchor ratings more difficult to interpret in relation to the other versions. For future studies, one possible consideration is to degrade the anchor using methods more parametrically related to the other sound quality versions to facilitate rating comparisons.

In this study, we identified high-frequency impairments as one contributing factor for the poor musical sound quality experienced by CI users. Future studies should address the possible musical sound quality benefits conferred by transmitting a greater amount of high-frequency information and increasing resolution within the higher-frequency ranges. Given the limited number of electrodes in current array designs, the device transmits frequencies within the range needed for speech perception. However, designing new arrays to increase the number of physical electrodes may allow the more basal electrodes to represent higher frequencies (>8 kHz) and smaller-frequency bandwidth assignments within the higher-frequency range (Hillman, Badi, Normann, Kertesz, & Shelton, 2003). Further application of strategies, such as current steering (i.e., virtual channels) and current focusing, may increase resolution within the transmitted frequencies and provide better musical sound quality (Berenstein, Mens, Mulder, & Vanpoucke, 2008; Koch, Downing, Osberger, & Litvak, 2007; Landsberger & Srinivasan, 2009).

Although traditional rating scale methods have allowed researchers to conclude that overall CI-mediated musical sound quality perception is poor, little attention has been focused on identifying which specific acoustic features of music contribute to these sound quality deficits (Gfeller et al., 2000; Lassaletta et al., 2008a). By providing experimental control over the type of degradation applied to the musical stimuli, CI-MUSHRA provides a unique tool for identifying and quantifying acoustic features that need to be more effectively transmitted by the CI device. In this study, we focused on high-frequency loss; however, CI-MUSHRA offers the potential to explore the impairments of a large range of acoustic parameters on musical sound quality for CI users.

Authors' Note

Declaration of Conflicting Interests

Funding

References

Berenstein, C. K., Mens, L. H., Mulder, J. J., & Vanpoucke, F. J. (2008). Current steering and current focusing in cochlear implants: Comparison of monopolar, tripolar, and virtual channel electrode configurations. *Ear & Hearing*, *29*(2), 250-260.

Brockmeier, S. J., Fitzgerald, D., Searle, O., Fitzgerald, H., Grasmeder, M., & Hilbig, S., et al. (2011). The MuSIC perception test: A novel battery for testing music perception of cochlear implant users. *Cochlear Implants International*, *12*(1), 10-20.

Driscoll, V. D., Oleson, J., Jiang, D., & Gfeller, K. (2009). Effects of training on recognition of musical instruments presented through cochlear implant simulations. *Journal of the American Academy of Audiology*, *20*(1), 71-82.

Galvin, J. J. III, Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169*, 518-533.

Gfeller, K., Christ, A., Knutson, J., Witt, S., & Mehr, M. (2003). The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal hearing adults. *Journal of Music Therapy*, *40*(2), 78-112.

Gfeller, K., Christ, A., Knutson, J. F., Witt, S., Murray, K. T., & Tyler, R. S. (2000). Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology*, *11*(7), 390-406.

Gfeller, K., Oleson, J., Knutson, J. F., Breheny, P., Driscoll, V., & Olszewski, C. (2008). Multivariate predictors of music perception and appraisal by adult cochlear implant users. *Journal of the American Academy of Audiology*, *19*(2), 120-134.

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., & Stordahl, J., et al. (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology*, *13*(3), 132-145.

Ginocchio, J. (2009). The effects of different amounts and types of music training on music style preference. *Bulletin of the Council for Research in Music Education*, *182*, 7-17.

Hillman, T., Badi, A. N., Normann, R. A., Kertesz, T., & Shelton, C. (2003). Cochlear nerve stimulation with a 3-dimensional penetrating electrode array. *Otology & Neurotology*, *24*(5), 764-768.

International Telecommunication Union. (2003). *Method for the subjective assessment of intermediate sound quality* (MUSHRA; ITU-R. Recommendation.BS.1534-1). Geneva, Switzerland: Author.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., & Nie, K., et al. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception test. *Ear & Hearing*, *30*(4), 411-418.

Kemp, A. E. (1996). *The musical temerament: Pyschology and personality of musicans*. New York, NY: Oxford University Press.

Koch, D. B., Downing, M., Osberger, M. J., & Litvak, L. (2007). Using current steering to increase spectral resolution in CII and HiRes 90K users. *Ear & Hearing*, *28*(Suppl. 2), 38-41.

Landsberger, D. M., & Srinivasan, A. G. (2009). Virtual channel discrimination is improved by current focusing in cochlear implant recipients. *Hearing Research*, *254*(1-2), 34-41.

Lassaletta, L., Castro, A., Bastarrica, M., Perez-Mora, R., Herran, B., & Sanz, L., et al. (2008a). Changes in listening habits and quality of musical sound after cochlear implantation. *Otolaryngology—Head & Neck Surgery*, *138*(3), 363-367.

Lassaletta, L., Castro, A., Bastarrica, M., Perez-Mora, R., Herran, B., & Sanz, L., et al. (2008b). Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringológica Española*, *59*(5), 228-234.

LeBlanc, A. (1982). An interactive theory of musical preference. *Journal of Music Theory*, *19*, 29-45.

LeBlanc, A. (1991, March). *Effect of maturation/aging on music listening preference: A review of the literature*. Paper presented at the 9th National Symposium, on Research in Musical Behavior, Cannon Beach, OR, USA.

Limb, C. J., & Rubinstein, J. T. (2012). Current research on music perception in cochlear implant users. *Otolaryngologic Clinics of North America*, *45*(1), 129-140.

Looi, V., Winter, P., Anderson, I., & Sucher, C. (2011). A music quality rating test battery for cochlear implant users to compare the FSP and HDCIS strategies for music appreciation. *International Journal of Audiology*, *50*(8), 503-518.

McDermott, H. J. (2004). Music perception with cochlear implants: a review. *Trends in Amplification*, *8*(2), 49-82.

Migirov, L., Kronenberg, J., & Henkin, Y. (2009). Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology, and Laryngology*, *118*(5), 350-355.

Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., & Hawthorne, M. (2003). Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear Implants International*, *4*(2), 85-95.

Nater, U. M., Abbruzzese, E., Kreb, M., & Elhert, U. (2006). Sex differerences in emotional and psychophysiological responses to musical stimuli. *International Journal of Psychophysiology*, *62*(6), 300-308.

Roy, A. T., Jiradejvong, P., Carver, C., & Limb, C. J. (2012). Assessment of sound quality perception in cochlear implant users during music listening. *Otology & Neurotology*, *33*(3), 319-327.

Schuessler, K. F. (1948). Social background and musical taste. *American Sociological Review*, *13*(3), 330-335.

Snow, W. B. (1931). Audible frequency ranges of music, speech and noise. *Journal of the Acoustical Society of America*, *3*, 616-627.

Vincent, E. (2005). *MUSHRAM: A MATLAB interface for MUSHRA listening tests*. Retrieved from http://www.elec.qmul.ac.uk/people/emmanuelv/mushram