EVALUATION OF AN INTERACTIVE MUSIC AWARENESS

PROGRAM FOR COCHLEAR IMPLANT RECIPIENTS

Rachel M. van Besouw, Benjamin R. Oliver, Mary L. Grasmeder,

Sarah M. Hodkinson and Heidi Solheim

University of Southampton, UK

Abstract

The objective of this study was to evaluate the efficacy of a prototype interactive music awareness program (IMAP) for adult cochlear implant (CI) users. An unblinded, randomized, crossover design was used. Twenty-one CI users were recruited and allocated to two groups. Group 1 received the IMAP first, followed by a retention of learning phase. Group 2 were given the IMAP after 12 weeks. Participants were instructed to undertake two half-hour sessions per week at home over 12 weeks. Both groups attended appointments at the start, halfway through, and at the end of the trial. At each appointment participants completed tests of speech perception, melodic contour identification and instrument recognition, rated the sound quality of music, and indicated their music listening habits. Sixteen participants completed the study. Following training both groups showed improved instrument recognition abilities and feedback suggests further positive impact on participants' lives. The findings suggest that the IMAP is beneficial for music perception and in particular, improved instrument recognition.

Key words:

cochlear implant, music perception, participatory design, training program, evaluation

Music perception is known to be challenging for many cochlear implant (CI) users due to the limitations of CI processing, particularly regarding the loss of temporal fine structure necessary for the accurate perception of pitch and timbre, and the way in which CIs interface with an already impaired auditory system (for comprehensive reviews, see McDermott, 2004; Looi, 2008; Looi, Gfeller, & Driscoll, 2012). The difficulties that face many CI users is borne out in their negative appraisal and sound quality ratings of music, with the timbre of instruments being described or rated as "thin", "shrill", "noisy", "tinny", "empty" or "confusing" (Gfeller, Christ, et al., 2000; Gfeller, Witt, Adamek, et al., 2002; Looi & She, 2010). As a consequence, postlingually deafened adult CI users report a decline in music listening habits following implantation, low satisfaction in listening to music and a desire to hear music as it sounded prior to their hearing loss or as it would be perceived with normal hearing (Gfeller, Christ, et al., 2000; Leal et al., 2003; Mirza, Douglas, Lindsay, Hildreth, & Hawthorne, 2003; Lassaletta et al., 2008; Looi & She, 2010; Looi et al., 2012; Philips et al., 2012).

A few studies have investigated the therapeutic value of music listening exercises on the ability of adult CI users to perceive and enjoy music with varying success. Gfeller, Witt, et al. (2000) investigated the effects of a music training program on melody recognition and appraisal in 11 adult CI users with a further 9 CI users as a control group using random assignment. The training comprised simple pitch, timbre and melody tasks as well as self-directed exploratory tasks. Pre- and post-training measures included simple melody recognition, complex song recognition and complex song appraisal. Gfeller and colleagues reported a significant interaction between the training and control groups from pre- to post-training for the recognition and appraisal of complex songs, with the training group achieving higher scores and giving more positive ratings. No change for either group was observed for simple melody recognition.

Gfeller, Witt, Woodworth, Mehr, and Knutson (2002) presented further evidence supporting the positive impact of formal training on timbre recognition and appraisal. Twenty-four postlingually deafened adult CI users were randomly assigned to either a control or training group. The training group underwent 12 weeks of computer-based training comprising 48 lessons, each lasting for about ten minutes while the control group received no intervention. Improvements in timbre recognition scores and appraisal ratings of musical instrument excerpts from pre- to post-test were reported for the training group, but not the control group; however, interaction effects, which would demonstrate that the improvements in the training group were statistically significantly greater than the improvements in the control group, were not reported.

Galvin, Fu, and Nogaki (2007) developed and evaluated a closed set melodic contour identification (MCI) test for CI users and then used this software to train a small group of adult CI users, with five participants training for 30 minutes per day over an indeterminate period and one participant training for three hours per day over five days. MCI performance (using a different frequency range to the MCI task used for training) was found to improve with training. Whilst the generalizability of this improvement might be questioned due to the similarity of the training and test materials, the authors also report improvement in post-training measures of familiar melody identification for four participants, suggesting that targeted MCI training benefits melody perception in general.

Driscoll (2012) investigated the effect of computer-based training specifically on instrument recognition with adult CI users. Participants were randomly assigned to one of three feedback conditions; correct/incorrect feedback, correct/incorrect feedback followed by the answer, and directed instruction on the instrument presented followed by correct/incorrect feedback. The study design did not include a control group. Over a period of five weeks, participants undertook 15 ten-minute sessions comprising recordings of eight instruments.

Performance improved from pre-test to week 3 and from week 3 to week 5 of the training irrespective of the feedback condition. However, participants were trained using the same stimuli used for testing, bringing into question the generalizability of the training effect reported.

Petersen, Mortensen, Hansen, and Vuust (2012) investigated the effect of one-to-one musical ear training on newly implanted adult CI users. Eighteen participants were assigned to either a training group or a no-intervention control group based on duration and degree of deafness, use of a contralateral hearing aid, and availability. Over a period of six months, participants in the training group received weekly 1-hour one-to-one musical ear training sessions with a professional music teacher, supported by computer-based training at home. Both groups were tested using a battery of eight speech and music perception measures at 'baseline' within 14 days of initial tuning of the CI, at three months and again at six months (the end of the trial). All participants in the training group completed the one-to-one sessions; however, the extent to which participants used the computer-based training program is unclear as this was not recorded. Statistically significant interactions between group and time were reported for measures of instrument recognition, MCI, and rhythm discrimination, although the significance level was not adjusted for multiple comparisons. Again, the computer-based training included instrument samples and melodic contours that also featured in the instrument recognition and MCI outcome measures.

A general limitation of the above studies is that they either had no control group, or had a control group that did not go on to receive the training. Levitt & List (2011) propose that a potential "Hawthorne effect" might arise through the "scrutiny and emphasis on process accompanying experimentation". In other words, simply reading the participant information sheet and instructions, signing the consent form and attending appointments might bias performance. If the same attention is given to a control group, any change in

performance due to this effect will affect both control and intervention groups similarly, with differences between the groups post-intervention being attributable to the intervention[1]. In this case, the efficacy of the intervention would be exaggerated in the absence of a control group. However, a control group that does not go on to receive the intervention might be less motivated to perform well and even exhibit resentful demoralization (Dunn et al. 2003), which would also exaggerate any improvement in performance shown by the intervention group. Additionally, familiarity with the test material could be a confounding factor. Nonetheless, the evidence that CI users are able to "relearn" the sounds of musical instruments is particularly compelling and is further supported by case studies in other music perception experiments. Fujita and Ito (1999) noted that CI users in their study were able to learn the timbres of five instruments that were presented repeatedly and, in a study on the timbre recognition abilities of postlingually deafened adult CI users and normal hearing (NH) listeners, Gfeller, Witt, Woodworth, et al. (2002) reported the case of one CI user who correctly identified all of the instruments in an instrument recognition task, outperforming several NH listeners. The authors report that this individual intensively relearned the sounds of various instruments following implantation, suggesting that with practice, timbre perception can be improved for some CI users.

Two recent music perception survey studies with postlingually deafened adult CI users indicate demand for music training as part of a rehabilitation program (Looi & She, 2010; Philips et al., 2012). Of the 40 CI users surveyed by Philips et al., 52% of respondents agreed that being able to enjoy music was important and 65% agreed that learning to listen to music during rehabilitation is useful. In the study by Looi & She, 45 out of 84 respondents indicated that they would be interested in undertaking a music training program. The majority

---

[1] Although the intervention itself could be a further source of Hawthorne effect; see Levitt & List (2011) for further discussion.

of participants indicated that they would be prepared to undertake 30 minute sessions 2-3

times per week and in terms of content, participants indicated that training should include a

wide range of musical styles and focus on their ability to recognize previously and commonly

known tunes.

Whilst the above studies suggest that music training might be beneficial, and despite

demand for music training programs, availability of music resources that have been

developed and evaluated with adult CI users is limited. The goal of programs that have been

used previously for research has also largely been to improve CI users' perception of certain

aspects of music, without necessarily addressing activity limitations arising from difficulties

in perceiving music through a CI, such as listening to a radio, appreciating the soundtrack of

a film or creating music.

In order to address users' needs and desires, Oliver and colleagues adopted a

participatory design approach in developing a prototype "Interactive Music Awareness

Program" (IMAP) with adult CI users (Oliver, van Besouw, & Nicholls, 2012; van Besouw,

Nicholls, Oliver, Hodkinson, & Grasmeder, 2014). In a series of initial consultations, adult CI

users indicated that they not only wanted to improve their music perception abilities, but also

wanted tools to help them (re)engage with music. Feedback from the consultations informed

the development of software applications that enabled users to interact with music in creative

ways. The applications were tested in a series of workshops with adult CI users, where they

were rated by attendees as being one of the two most useful or interesting aspects (van

Besouw et al., 2014), and then refined and incorporated into the prototype IMAP.

The primary research objective of the trial discussed in this study was to evaluate the

efficacy of the prototype IMAP, with the specific aims of determining if the program is

beneficial for music perception, if this benefit generalizes to speech in noise (SIN) perception

and if the program has a positive impact on ratings of sound quality and music listening

habits. On the basis of the research reviewed, it was hypothesized that MCI, instrument recognition and ratings of music sound quality would improve following use of the program, and that participants would choose to listen to music more frequently as a result of the program. In addition, it was hypothesized that music-based auditory training would result in perceptual learning that would generalize to SIN perception. Sentences, in particular, test cognitive skills including attention and working memory as well as basic auditory function (McArdle, Wilson, & Burks, 2005); skills relevant to music listening and which appear to contribute to musicians' better performance on SIN tasks (Parbery-Clark, Skoe, Lam, & Kraus, 2009). Evidence of neuroplasticity concomitant with improved SIN performance has also been reported for short-term auditory training (Russo, Nicol, Zecker, Hayes, & Kraus, 2005; de Boer & Thornton, 2008), although this has yet to be shown for music-based auditory training for listeners with normal or impaired hearing.

A secondary objective for this study was to obtain feedback on the prototype IMAP in order to improve it. Findings for the primary research objective are presented here.

Materials and Methods

Program

In the interest of brevity, a short description of the prototype IMAP is given here; for a more detailed description see Oliver et al. (2012). The prototype IMAP was housed on a USB memory stick and could be used on Windows or Apple platforms. The user was guided through 24 half-hour sessions by means of an html-template that linked to interactive Max/MSP[2] standalone applications, enabling users to create and manipulate music. Applications included graphical mixers that allowed users to control the instrumental/vocal

---

[2] Max/MSP is a visual programming language by developer, Cycling '74, for working with multimedia. Full documentation available at: https://cycling74.com/products/max/

mix, pitch and speed of songs, music players that allowed users to select different melodies and instrument combinations, a drum machine with samples of environmental sounds that could be looped and layered, enabling users to compose new rhythmic structures, and an audio and video player that enabled users to combine audio and video clips to explore how music contributes to the mood, meaning and aesthetics of a film. The applications were introduced through text and video instructions. Each odd numbered session concluded with a directed online listening task, ultimately encouraging users to discover music on sites such as YouTube. Each even numbered session concluded with a music perception test, which informally tested users' abilities in pulse detection, melodic contour perception and instrument recognition (Oliver et al., 2012).

For the trial each participant was given a pamphlet with additional instructions, a troubleshooting guide, tips on listening to audio on a computer using loudspeakers, headphones or direct connection with an isolation cable, and space to make observations regarding their own progress. Participants were allowed to choose the listening format that they would normally use when listening to music on their home computer. At the end of each session in the IMAP participants were prompted to complete a password protected online survey, where they were asked to enter how long they had spent on the session, rate the session, rate the software, and give optional free-response feedback.

Participants

The study was approved by the National Research Ethics Service (reference 11/SC/0436), the Institute of Sound and Vibration Research Human Experimentation Safety and Ethics Committee (reference 1250), and the University of Southampton Research Governance Office (reference RGO 8306).

At the time of the study 148 adults at the University of Southampton Auditory Implant Service met the inclusion criteria and were sent a study invitation. Participants were included if they: had and were able to use a Windows PC or Mac, were able to travel to the University of Southampton for the music perception appointments, had achieved ≥60% on the Bench, Kowal, & Bamford (1979) Bamford-Kowal-Bench (BKB) sentence test in quiet at their most recent assessment, had been using their implant for at least 1 year, were capable of giving informed consent, and were not involved in other studies that could confound the results.

An unblinded, randomized, crossover design was used. Twenty-one participants (14% of adults who received the study invitation: 11 female, 10 male) were recruited and randomly allocated to two groups; group 1 (11 participants) and group 2 (10 participants) using a pseudo-random number generator in MATLAB to generate a blocked randomization list. The decision to use this form of allocation was based on the need to prevent potential conscious or subconscious bias by the researcher, and overlap of the recruitment phase with the start of the trial.

Group 1 received the program first, followed by a 12-week retention of learning phase. Group 2 were given no intervention for the first 12 weeks (the control phase) and then were given the program to use, with no subsequent retention of learning phase (Figure 1). In their respective training phases both groups were instructed to undertake two, half-hour sessions per week at home over 12 weeks. Both groups were required to attend three 2-hour music perception assessment appointments at the start (T1), after 12 weeks (T2) and after 24 weeks (T3). In the appointment prior to the training phase (at T1 for group 1 and T2 for group 2), participants were shown how to access and use the IMAP on a computer. Participants were then emailed by the experimenter every four weeks of the training phase to check that they were not experiencing technical difficulties.

During the training phase two participants (one from each group) withdrew due to the time commitment involved, one withdrew from group 1 due to difficulties using a computer and one participant withdrew from group 2 a week before the final appointment at T3 due to bereavement. Between T1 and T2 a further participant from group 2 (control phase) withdrew without giving a reason, leaving 16 participants in total; nine in group 1 and seven in group 2. Participant characteristics for these 16 are given in Table 1.

In the UK unilateral cochlear implantation is recommended for adults with severe-to-profound bilateral cochlear hearing loss; hence the most of the participants recruited were unilateral CI users. Three participants had some residual low frequency hearing in the ear contralateral to the CI and therefore also used a hearing aid (known as bimodal stimulation). Funding is not normally available for adults to be implanted bilaterally. However, if funding is made available from another source, bilateral implantation may be offered, as in the case of P17.

[Figure 1 here]

[Table 1 here]

Outcome Measures

At each appointment participants were asked to undertake tests of speech perception, MCI and instrument recognition, rate five pieces of music in terms of sound quality, and indicate their music listening habits. The tests were always presented in this order to ensure that effects of practice, boredom, and fatigue would be similar across appointments. Care was taken to ensure that the stimuli used for each outcome measure were unique to that measure and did not feature in the prototype IMAP to ensure that participants were not merely training to do the tests; for example, the melodic contours used in the informal music perception test in the prototype IMAP were excerpts of real-world monophonic piano melodies, with a graphic symbol that traced out the shape of the contour as the contour played. Hence the

contours differed to those used for the MCI outcome measure in terms of both the contour and timbre[3]. Likewise, the informal test of instrument recognition in the IMAP used excerpts of recordings of 16 instruments, which differed from the samples used in the instrument recognition outcome measure.

The participants performed all tests using their usual listening mode (unilateral, bilateral or bimodal) and program setting to avoid imposing a mode of listening that they might be unaccustomed to. All stimuli were presented from a laptop via a Behringer UCA202 audio interface over a single Fostex 6301B loudspeaker positioned 1.5 m at 0º azimuth at head height in front of the participant. Stimuli were presented at 65-70 dB(A) as measured at the participant's ear using a calibrated Kamplex KM4 (IEC 651 Type 2) sound level meter. Levels were checked prior to each test. All stimuli were 44.1 kHz 16-bit wav files with the exception of the stimuli used to assess music appreciation; these were MP3 files encoded with a 256 kbit/s variable bit rate.

Speech in noise (SIN) perception. SIN perception was assessed using the closed set UK-matrix sentence test, an English version of the Hagerman, Oldenburg and Dantale II tests (Hagerman, 1982; Wagener, Josvassen, & Ardenkjær, 2003). Each sentence in the UK-matrix test contains five words and the syntactical structure of each sentence is identical (name, verb, numeral, adjective, object). However, the sentences are semantically neutral and therefore have low predictability. Following presentation of a sentence, participants are required to select one of 10 alternatives for each part of the sentence in a 10-by-5 matrix (Figure 2).

---

[3] Following the trial, the informal melodic contour test in the IMAP was made more similar to the MCI test in response to feedback from participants who found it difficult to follow the moving graphic symbol.

The sentences were spoken by a female talker and presented at a fixed level of 65 dB(A). Speech spectrum-shaped noise was initially set to 30 dB below the level of the speech and altered for subsequent sentences using

$$\Delta L = -\frac{prev - tar}{slope}$$ (Hagerman & Kinefors, 1995; Brand & Kollmeier, 2002) (1)

where ΔL is the change in noise level in dB, *prev* is the discrimination value obtained in the previous sentence (a score out of 5 words correct), *tar* is the target discrimination value and *slope* is the estimated slope of the discrimination function. For this study *slope* = 0.2 dB^{-1} and *tar* = 0.8 (i.e. 80% intelligibility). The speech reception threshold (SRT) in noise (defined as a signal-to-noise ratio in dB) was calculated using the average of the final six of eight reversals in noise level.

Participants were required to click a button to elicit each sentence and then select five words in the matrix (one response per column) and were instructed to guess if unsure of the answer. Prior to the test participants were presented with three practice sentences in order to familiarize them with stimuli and procedure.

[Figure 2 here]

MCI. A modified version of the MCI task reported by Galvin et al. (2007) was used to assess the ability of participants to follow a melody. The stimuli comprised nine melodic contours (Rising, Rising-Flat, Rising-Falling, Flat-Rising, Flat, Flat-Falling, Falling-Rising, Falling-Flat, and Falling), each with five notes of 250 ms duration. The notes where synthesized harmonic complexes containing the fundamental frequency (f_0), $2f_0$ at -3dB, and $3f_0$ at -6dB. Raised cosine onset and offset ramps of 10 ms were applied to each note and the interval between notes was 50 ms. The short-term loudness of each note was adjusted to 65

phon using the Glasberg and Moore (2002) model of time-varying loudness implemented in the Genesis Loudness Toolbox for MATLAB (Genesis, 2009).

Interval sizes between successive notes of 5, 4, 3, 2, and 1 semitones were used with a root note of A4, such that for an interval size of 5 semitones, the Rising contour spanned notes A4 to F6 (1397 Hz), and for an interval size of 1 semitone, the Rising contour spanned notes A4 to D5 (587 Hz). Unlike the MCI task reported by Galvin et al., only one root note was used (note A4 at 440 Hz) and intervals between successive notes were not randomized. Instead, the stimulus set was blocked according to interval size, making the test gradually more difficult by reducing the interval size from 5 semitones down to 1 semitone. Each block comprised three presentations of each of the nine contours presented in a random order, giving a total of 135 trials and taking participants 19 minutes on average to complete.

Participants were required to click a button to elicit each stimulus and then click on one of nine representations of the contours to indicate their response (Figure 3). They were able to repeat each stimulus once only and were instructed to guess if unsure of the answer. No feedback was given. Prior to the test participants were twice presented with each of the contours in turn using an interval size of 5 semitones in order to familiarize them with the stimuli and the contour representations.

[Figure 3 here]

Instrument recognition. Instrument recognition was tested using an 8-alternative forced choice procedure, using samples of a bowed violin, plucked acoustic guitar, trumpet, French horn, piano, xylophone, flute, and oboe. All instrument samples were generated using the Vienna Symphonic Library of orchestral samples in Logic Pro 7, with the exception of the bowed violin and xylophone, which were generated using the EXS24 Sampler in Logic Pro 7 and the DSLMusicDevice in Sibelius 5.1 respectively. Each stimulus comprised a 7 note

ascending and descending staccato arpeggio in the key of C major, starting and ending on

note C4 (261.6 Hz) at a rate of 110 notes per minute. Note length varied from 0.2 to 0.4 s

depending on the temporal envelope of the instrument sample. As with the stimuli for the

MCI test, the short-term loudness of each note was adjusted to 65 phon.

Each instrument was presented three times in a random order, giving a total of 24

trials and taking participants 5 minutes on average to complete. Participants were required to

click a button to elicit each stimulus and then click on one of eight representations of the

instruments to indicate their response (Figure 4). They were able to repeat each stimulus once

only and were instructed to guess if unsure of the answer. No feedback was given. Prior to

the test participants were twice presented with each of the instruments in turn in order to

familiarize them with the stimuli and representations.

[Figure 4 here]

Music sound quality. Music sound quality was evaluated using the visual analogue

scales (VAS) for pleasantness (unpleasant-pleasant) and naturalness (unnatural-natural)

described by Looi, Winter, Anderson, and Sucher (2011), and a third VAS for clearness with

the descriptors unclear-clear. The three scales were presented on a screen with the two

contrasting adjectives equidistant from the center. The descriptors unpleasant, unnatural, and

unclear were positioned to the far left of the scale, with their opposites to the far right. The

default starting position of the sliders for each scale were to the far left and participants were

required to adjust these as in the study of Looi et al.

Ratings were made for five tracks of unfamiliar music representative of the genres:

blues, classical, country, jazz, and pop (Table 2). Although Looi et al. did not observe an

effect of song familiarity on sound quality ratings in their study, other studies have reported

exposure effects on liking (see for example Peretz, Gaudreau, & Bonnel, 1998), which could

bias ratings of sound quality. Familiarity with the tracks was therefore established at T1 by

asking participants if they recognized any of the artists and track titles listed in Table 2 prior

to and immediately following the ratings. The five tracks were reported as being unfamiliar

by all of the participants. CI users find music recognition difficult, particularly for music that

they were unfamiliar with prior to deafness (Gfeller et al. 2005). It was therefore considered

unlikely that the brief exposure to the tracks T1 would influence ratings at T2 or at T3.

Participants played each track from the start and had the option to stop and play the

track from the beginning again. Whilst listening to the music, participants were able to adjust

the sliders until satisfied with their ratings and could choose to listen to each track as long as

they wished. Participants were encouraged to rate the sound quality regardless of whether or

not they liked the music.

[Table 2 here]

Music listening habits. At T1 participants were asked to indicate how often they

currently chose to listen to music and were given the options: daily, weekly, monthly, less

than monthly, and never. At subsequent appointments participants were asked how often they

had chosen to listen to music since their last appointment and were given the same options. If

a participant had just completed the training phase, they were asked to consider how often

they had chosen to listen to music in addition to the music that they had been exposed to in

the program sessions.

Statistical Analysis

For normally distributed data according to the Shapiro-Wilk test, independent-

samples *t*-tests (*t*) were used to compare the groups at T1, mixed analysis of variance

(ANOVA) tests (*F*) were used to explore the group x session interaction from T1 to T2, paired-samples *t*-tests were used to compare scores between sessions within each of the groups, and the Pearson's correlation coefficient (*r*) to explore relationships between variables.

Where the data were non-normally distributed according to the Shapiro-Wilk test, non-parametric tests were used (the Wilcoxon-signed rank test (*W*) for two related samples, the Mann-Whitney test (*U*) for two independent samples, and the Kendall's tau (*τ*) correlation coefficient). To explore the group x session interaction from T1 to T2 a Mann-Whitney test was used to compare change in performance between the groups.

For each of the five outcome measures (UK-matrix SRT, MCI, instrument recognition, music sound quality, and music listening habits) a Bonferroni corrected significance level of .01 was applied when exploring the group x session interaction from T1 to T2 and also change in performance for group 2 between T2 and T3. One-tailed test values are reported where prior directional hypotheses were made.

Throughout section 3 the median (*Mdn*) and interquartile range (*IQR*) are reported to aid comparison with conditions for which the data are not normally distributed.

Results

At T1, there were differences between the groups in terms of their characteristics. In group 2, five of the participants had undertaken formal music training or study prior to implantation, whereas none of group 1 had such experience. Duration of implant use for group 1 (*Mdn* = 77 months, *IQR* = 70) did not differ statistically significantly from group 2 (*Mdn* = 47 months, *IQR* = 28), *U* = 26.0, *z* = -0.58, *p* = .585 (2-tailed) and likewise, age for group 1 (*Mdn* = 55 years, *IQR* = 24) did not differ significantly from group 2 (*Mdn* = 60 years, *IQR* = 16), *t*(14) = -0.21, *p* = .837 (2-tailed). Group 1's BKB in quiet scores (*Mdn* =

94%, *IQR* = 13) were not statistically significantly different from group 2's scores (*Mdn* = 99%, *IQR* = 16), *U* = 24.5, *z* = -0.75, *p* = .477 (2-tailed), although many participants were performing at or close to ceiling for this test (see Table 1), which might have masked performance differences between the groups.

SIN Perception

At T1, the UK-matrix SRT for group 1 (*Mdn* = 17.9 dB, *IQR* = 27.5) did not differ significantly from group 2 (*Mdn* = 1.5 dB, *IQR* = 40.7), *U* = 18.5, *z* = -1.38, *p* = .183 (2-tailed). At T2, the UK-matrix SRT for group 1 (who had just completed the IMAP) were significantly better (*Mdn* = 4.7 dB, *IQR* = 27.9) than at T1, *W* = 1.0, *z* = -2.55, *p* = .004, *r* = -.60 (1-tailed), whereas group 2's scores (*Mdn* = -0.5 dB, *IQR* = 38.3) did not differ significantly, *W* = 7.0, *z* = -1.18, *p* = .148 (1-tailed). Comparison of the change in UK-matrix SRT from T1 to T2 between the groups approached statistical significance, *U* = 15.5, *z* = -1.70. *p* = .048, *r* = -.43 (1-tailed, using a significance level of .01) with group 1 (who had just completed the IMAP) showing greater improvement (Figure 5). However, no change in UK-matrix SRT between T2 and T3 was observed for group 2 following use of the IMAP (*Mdn* = 0.2 dB, *IQR* = 31.8), *W* = 12.0, *z* = -0.34, *p* = .406 (1-tailed). Group 1's scores between T2 and T3 (the retention of learning phase) did not differ significantly (*Mdn* = 4.6 dB, *IQR* = 26.7), *W* = 17.0, *z* = -0.14, *p* = .480 (1-tailed).

MCI

At T1, scores of 100% were achieved by four participants in group 2 and two participants in group 1 for at least one of the interval conditions tested. For each participant, mean performance was calculated across the five interval conditions of the MCI test. At T1, MCI performance for group 1 (*Mdn* = 54.1%, *IQR* = 31.9) was statistically significantly

lower than for group 2 (*Mdn* = 94.8%, *IQR* = 28.9), *t*(14) = -2.21, *p* = .044, *r* = .51 (2-tailed).

At T2, MCI performance increased significantly for group 1 (*Mdn* = 77.0%, *IQR* = 32.3), *t*(8)

= 4.32, *p* = .001, *r* = .84 (1-tailed), with all 9 participants in this group achieving higher

scores following use of the IMAP. Despite a slight decrease in the median score, there was

also a statistically significant improvement in MCI performance following the control phase

for group 2 (*Mdn* = 92.6%, *IQR* = 26.7), *W* = 3.0, *z* = -1.86, *p* = .039, *r* = -.50 (1-tailed) with

6 out of 7 participants achieving higher scores (Figure 5).

Comparison of the change in MCI performance from T1 to T2 between the groups

approached statistical significance, *U* = 41.0, *z* = -1.96, *p* = .027, *r* = -.49 (1-tailed, using a

significance level of .01), with group 1 appearing to show greater improvement; however, at

T2, scores of 100% were achieved by five participants in group 2 and three participants in

group 1 for at least one of the interval conditions tested, including one participant in group 1

who achieved 100% for the 1 semitone interval condition at T2 and again at T3. Whilst no

participant achieved 100% at all interval levels of the MCI test, the ceiling effects

experienced by participants for many of the interval levels of the test would have limited the

degree of improvement that could be seen.

No change in MCI performance between T2 and T3 was observed for group 2

following use of the IMAP (*Mdn* = 91.1%, *IQR* = 25.3), *W* = 7.0, *z* = -1.18, *p* = .148

(1-tailed), or for group 1 following the retention of learning phase (*Mdn* = 69.6%, *IQR* =

33.3), *t*(8) = 1.62, *p* = .072 (1-tailed). At T3, scores of 100% were achieved by five

participants in group 2 and two participants in group 1 for at least one of the interval

conditions tested. Again, any improvement in performance in group 2 due to using the IMAP

between T2 and T3 could not be fully evaluated due to ceiling effects.

Instrument Recognition

At T1, instrument recognition scores for group 1 (*Mdn* = 41.7%, *IQR* = 18.7) did not differ significantly from group 2 (*Mdn* = 54.2%, *IQR* = 33.3), $t(14) = -0.69$, $p = .502$ (2-tailed). A mixed ANOVA comparing the change in scores from time T1 to T2 for each of the groups indicated a significant main effect of time, $F(1, 14) = 5.40$, $p = .036$, $r = .53$. Post-hoc *t*-tests showed a 8.3 percentage points gain in Group 1 (*Mdn* = 50.0%, *IQR* = 29.2), $t(8) = 2.93$, $p = .010$, $r = .72$ (one-tailed). However, the interaction effect between time and group was not statistically significant, $F(1, 14) = 3.33$, $p = .089$ (using a significance level of .01). At T3, instrument recognition scores for group 2 following use of the IMAP were significantly higher (*Mdn* = 62.5%, *IQR* = 33.3) than at T2 (*Mdn* = 50.0%, *IQR* = 29.2), $t(6) = 2.10$, $p = .040$, $r = .65$ (1-tailed). Between T2 and T3, group 1's scores did not differ significantly (*Mdn* = 50.0%, *IQR* = 25.0), $t(8) = -1.24$, $p = .125$ (1-tailed), suggesting some retention of learning (Figure 5).

[Figure 5 here]

Relationship Between Speech and Music Perception Measures at T1

The pattern of results between the groups at T1 suggests a general relationship between UK-matrix SRT, MCI performance and timbre recognition, with better performers (group 2) performing well across these measures. To explore this, correlations were calculated between the UK-matrix SRT and MCI scores, the UK-matrix SRT and instrument recognition scores and the instrument recognition and MCI scores for all 16 participants. UK-matrix SRT was statistically significantly correlated with MCI ($\tau = -.38$, $p = .021$) and with instrument recognition ($\tau = -.47$, $p = .007$).

Music Sound Quality

For each participant, mean ratings for pleasantness, naturalness, and clearness were calculated across the five music tracks. Participants' mean ratings for pleasantness, naturalness, and clearness were found to be highly correlated at T1 (all $r \geq .84$, $p < .001$), T2 (all $r \geq .92$, $p < .001$) and T3 (all $r \geq .87$, $p < .001$), and therefore the pleasantness, naturalness and clearness ratings were averaged to produce a single measure of sound quality from 0 (low quality) to 100 (high quality).

At T1, sound quality ratings for group 1 (*Mdn* = 54.1, *IQR* = 41.6) did not differ significantly from group 2 (*Mdn* = 67.1, *IQR* = 35.0), $t(14)$ = -1.54, p = .147 (2-tailed). A mixed ANOVA comparing the change in ratings from time T1 to T2 for each of the groups indicated no significant effect of time, $F(1, 14)$ = 2.62, p = .128, and no interaction $F(1, 14)$ = 2.72, p = .121.

No change in sound quality ratings between T2 (*Mdn* = 62.1, *IQR* = 16.7) and T3 (*Mdn* = 63.4, *IQR* = 14.5) was observed for group 2 following use of the IMAP, $t(6)$ = 0.68, p = .261 (1-tailed). Group 1's ratings between T2 (*Mdn* = 64.5, *IQR* = 37.8) and T3 (*Mdn* = 62.1, *IQR* = 42.8) did not differ significantly, $t(8)$ = 1.28, p = .118 (1-tailed) (Figure 5).

Music Listening Habits

Numeric values were assigned to each of the categories daily (5), weekly (4), monthly (3), less than monthly (2), and never (1). At T1 the difference between self-reported music listening habits for group 1 (*Mdn* = 4, *IQR* = 4) and group 2 (*Mdn* = 5, *IQR* = 1) approached statistical significance, U = 15.5, z = -1.83, p = .079 (2-tailed), with group 1 listening to music less often.

Comparison of change in music listening habits from T1 to T2 between the groups was not statistically significant, U = 30.0, z = -0.18, p = .453 (1-tailed). At T2 and T3, all

participants in group 2 reported listening to music daily. At T3, six out of nine participants in

group 1 reported listening to music daily (*Mdn* = 5, *IQR* = 1), with no significant change from

T2 (*Mdn* = 4, *IQR* = 3), *W* = 0, *z* = -1.89, *p* = .063 (1-tailed) (Figure 6).

[Figure 6 here]

Correlation with Time Spent Using the IMAP

Ten participants completed all 24 IMAP sessions, five completed 19-23 sessions, and

one completed 10 sessions. The average time spent using the IMAP across the 16 participants

was 911.3 minutes ±246.4 SD (Figure 7). The average time spent per session was 41.2

minutes ±6.1 SD.

The study was not designed to explore the relationship between time spent on training

and performance, and the small sample size may explain why no statistically significant

correlations were observed between time (in minutes) spent using the IMAP and change in

scores from pre- to post-IMAP use for the UK-matrix test (*τ* = .25, *p* = .088), MCI (*r* = .02,

p = .466), instrument recognition (*r* = .02, *p* = .479), and ratings of sound quality (*r* = -.30,

p = .134). Likewise, no statistically significant correlations were observed between the

number of IMAP sessions completed and change in scores from pre- to post-IMAP use (all

p > .05).

[Figure 7 here]

Association between Improvement and Duration of Implant Use

CI user performance continues to improve up to 18-30 months post implantation and

for some individuals acclimatization continues to occur for up to 5 years (Tyler, Parkinson,

Woodworth, Lower, & Gantz, 1997). Duration of implant use for the participants in this

study ranged from 18–94 months and so it is feasible that improvements in performance for some individuals were due to acclimatization post-implantation. Duration of implant use was therefore correlated with each of the performance measures with the assumption that more recently implanted CI users would show greater change in performance. No statistically significant correlations were observed between duration of implant use and change in scores from pre- to post-IMAP use for the UK-matrix test ($\tau = .14$, $p = .222$), MCI ($r = .03$, $p = .450$) and instrument recognition ($r = .36$, $p = .088$).

Association between Improvement and Initial Level of Performance

Following the method described in Stacey et al. (2010) correlation analyses were used to explore the potential relationship between participants' level of performance and the extent to which they improved following IMAP use. No statistically significant correlations were observed between level of performance and extent of improvement for the UK-matrix test ($\tau = -.17$, $p = .184$) or instrument recognition ($r = .18$, $p = .259$). A moderate, negative correlation was observed for MCI ($r = -.49$, $p = .027$), suggesting greater improvement for poorer performers. However, this might be explained in part by the ceiling effects that limited the improvement that could be observed for the better performers.

Discussion

SIN Perception

Following training with the IMAP, the UK-matrix SRT for group 1 improved, whereas the SRT for group 2 did not. Although at T1 the UK-matrix SRT for group 1did not differ statistically significantly from group 2, group 2's scores at T1 are better than those for group 1 for all performance measures. It is therefore tempting to ascribe the improvement in group 1's UK matrix SRT post-training to different levels of prior performance between the

groups. However, the lack of correlation between level of performance and extent of improvement for the UK-matrix test does not support this. An alternative explanation for the improvement in SRT for group 1 at T2 could be prior exposure of the test at T1, although it can be argued that a similar change in SRT at T2 should be evident for group 2, which is not.

A number of participants commented that they found it difficult to remember the five words in each sentence and frequently reported recency and primacy effects; phenomena characteristic of sentence tests (McArdle et al., 2005). Two participants in each of the groups (P6, P10, P12, and P14) achieved a SRT of ≥30 dB at T1, T2 and T3 and were effectively performing the test in quiet, hence the large range in SRT values. In these cases, the SRT measure is likely to be more indicative of their ability to remember and recall words rather than a "reception threshold" per se. It is also noteworthy that P14 (who is pre-lingually deaf) and P10 (whose onset of deafness is uncertain) had the lowest BKB sentence scores of the cohort. It is thus possible that the change in performance seen for group 1 post training, but not observed for group 2 reflects differences in cognitive abilities between the groups that might have existed at T1, but were not measured.

MCI

At T1, MCI performance for group 2 was statistically significantly better than for group 1. Formal music training at an advanced level has been shown to be a predictive of CI users' abilities to pitch rank, recognize instruments, and recognize instrumental versions of familiar melodies (Gfeller et al., 2008), which might partly explain the difference in performance as five out of the seven members of group 2 had received prior formal music training (Table 1).

Both groups improved in MCI performance from T1 to T2, demonstrating practice effects. Comparison of the change in performance from T1 to T2 between the groups

approached statistical significance, with group 1 appearing to show more improvement than group 2 (Figure 5). Whilst this could be attributed to use of the IMAP, it is also likely that the ceiling effects experienced by participants for some of the MCI interval conditions limited the degree of improvement that could be measured, thus exaggerating this trend. No further improvement between T2 and T3 was observed for group 2, but again, this might have been limited by ceiling effects.

The high performance of participants on this test was unexpected. None of the participants in the study of Galvin et al. (2007) achieved 100% for any of the interval conditions in contrast to the present study, where one participant even achieved 100% for the 1 semitone interval condition. However, in the current study only one root note was presented and stimuli were blocked according to interval size, whereas Galvin et al. used three root notes and randomized presentation of the entire stimulus set, which would have made the test more difficult.

The practice effect seen for group 2 between T1 and T2 could potentially have been reduced had the MCI been administered repeatedly at, or prior to, T1 to achieve a reliable measure of baseline performance. However, the MCI itself has been used as an auditory training tool and it has been shown that MCI performance continues to improve with daily training for over a month (Galvin et al., 2007). Thus, the practice effects associated with procedural knowledge of the MCI are confounded with the development of auditory and cognitive skills beneficial for music. In this study, the inclusion of a no-treatment control phase at least provides some indication of the extent of the practice effects.

Instrument Recognition

Although the interaction between groups 1 and 2 from T1 to T2 does not reach statistical significance (which is likely due to sample size), both groups showed improvement

for the instrument recognition task following their respective training phases (Figure 5), a

trend that suggests a positive effect of training on instrument recognition. This is consistent

with the findings of Fujita and Ito (1999), Gfeller Witt, Adamek, et al. (2002), Gfeller, Witt,

Woodworth, et al. (2002), Petersen et al. (2012), and Driscoll (2012) who propose that

training provides CI users with the opportunity to acclimatize to musical stimuli that they

might otherwise not have persevered with due to the way in which music sounds post-

implantation. That is, with practice, CI users can relearn the timbres of musical instruments

as they now sound through their implant.

Music Sound Quality

Gfeller, Witt, Adamek, et al. (2002) reported improved timbre recognition and timbre

appraisal in CI users following 12 weeks of computer-based training, and proposed that

training "may facilitate acclimatization to the various sound qualities of instruments" and

alter "expectations about what constitutes aesthetic beauty". In the current study

improvement in instrument recognition was not mirrored in the sound quality ratings of five

pieces of music; however, participants were not rating the timbres of individual instruments.

Their judgments of sound quality are therefore likely to reflect different aspects of their

music perception, for example, their abilities to: recognize instruments in the mix, follow the

melody, and/or understand the lyrics (if present). That participants were able to rate sound

quality based on such aspects, either individually or in combination, might explain the

considerable range in ratings and lack of discernible improvement post-training.

At T1 sound quality ratings for group 2 were slightly higher than for group 1.

Although not statistically significant, this trend might be due the differences in current music

listening habits and prior formal music training between the groups, in line with the findings

of Gfeller et al. (2008), who reported higher musical appraisal ratings for CI users with more

music listening experience prior to implantation, more music experience post implantation, and better SIN perception. Prior formal music training could also account for the consistency of ratings for group 2 compared to those of group 1, where there is considerable spread.

Music Listening Habits

The music listening habits of groups 1 and 2 increased over the course of the trial (Figure 6), but no statistically significant increase in listening habits was observed for either group following training. The general increase in listening habits for both groups from T1 to T3 could be due to participation in a study that drew the participants' attention to musical stimuli. The 5 point scale used for evaluating listening habits had previously been used successfully to evaluate longer term changes in listening habits following a series of music workshops (van Besouw et al., 2014). However, in the present study, the options "monthly" and "less than monthly" were not particularly informative as the appointments were only 1 month apart. A frequency of listening scale with response categories independent of the time span between appointments (e.g. never, very rarely, rarely, occasionally, frequently, very frequently) might have been more sensitive to changes in listening habits.

A limitation of frequency measures of music listening habits is that they do not capture individuals' experiences and feelings in the way that a measure of music-related quality of life might. At the end of the trial participants gave feedback on the IMAP and comments relating to their experiences include:

"It [the training] made me listen to music that I haven't listened to for years." (P5)

"This programme has made me listen to music and appreciate differences in sounds that I had not heard before (e.g. the violin)." (P7)

"The training made me think more about the music and listen more carefully - I am

getting more out of it." (P13)

"Since the training I have been listening to Classic FM in the car - I would never have

done this before. I now make a point of doing this…" (P15)

The feedback above suggests the need for a measure sensitive to changes in what music

participants listen to, where, when, and why they listen to music, and their interest in and

awareness of music, in order to more fully evaluate the impact of training programs like the

IMAP.

Correlation with Time Spent Training

Participants were instructed to spend about 30 minutes per session in the IMAP and

undertake two sessions per week for a period of 12 weeks. Over the course of the IMAP

participants were expected to spend ~720 minutes using it. The average time spent per

session was 41.2 minutes; longer than anticipated, which could be due to the nature of the

open-ended activities as well as time spent navigating and loading applications. No

statistically significant correlations were observed between the outcome measures and time

spent using the IMAP or the number of sessions completed, which could be due to the small

sample size and differences in computer confidence, with less confident computer users

taking longer to undertake the activities. This might explain why time spent on the IMAP did

not always reflect the number of sessions completed. For example, P18 spent the least

amount of time, 475 minutes, using the IMAP, but completed 19 sessions, whereas P4

completed only 10 sessions, but spent 650 minutes using the IMAP. Future trials of this and

similar computer-based aural rehabilitation resources could include a measure of computer confidence.

Attrition

Excluding the exceptional circumstance of bereavement, four out of the 21 participants recruited withdrew from the study due to the time commitment involved, difficulties using a computer, and an unknown reason. In comparison, none of the nine participants withdrew from the one-to-one musical ear training sessions in the study of Petersen et al. (2012). Advantages of web-based over one-to-one training sessions include cost, travel and the convenience of being able to undertake sessions at leisure. However, the high attrition rate of web-based interventions is reported to be "one of the fundamental characteristics and methodological challenges in the evaluation of eHealth applications" (Eysenbach, 2005), and could be due in part to a desire for face-to-face support; at the end of the trial one participant commented "More contact with instructors would be a help" (P4), and another suggested having "an annual workshop to review about music" (P10). Interestingly, the two participants who withdrew due to the time commitment involved were the last two participants recruited. Eysenbach (2005) proposed that hesitation to participate in an eHealth trial "may be an early indicator for a potential dropout".

Differences in Group Characteristics and Potential Differences in Group Behavior

Prior to T1, the only performance measure with which the groups could be compared was their BKB in quiet scores and these were not statistically significantly different, potentially due to ceiling effects. Likewise, UK-matrix SRT at T1 did not differ significantly between the groups; yet from Figure 5 is it apparent that group 2's scores at T1 are better than those for group 1 for all performance measures and this is particularly clear for MCI where

the difference between the groups is statistically significant. This suggests that, on the whole, group 2 were better performers, which could be related to (1) the known difference in prior formal music training between the groups, or (2) unmeasured differences in their pitch acuity and/or cognitive processing abilities. The case for (1) is mixed; Gfeller et al. (2005) state that "in studies of timbre recognition, timbre appraisal, pitch discrimination, and melody recognition, musical training before implantation is not predictive of implant benefit"; however, in a later study Gfeller et al. (2008) found prior music training to be a significant predictor of pitch ranking ability, melody recognition and instrument recognition. In support of (2), Gfeller et al. (2008) also found cognitive factors to be predictive of pitch ranking ability and melody recognition. Furthermore, Gfeller et al. (2007) observed correlations between melody recognition, pitch ranking ability and SIN perception, and suggested that SIN perception is related to pitch acuity, which is supported by the statistically significant correlation between SIN perception and MCI observed in the present study.

A third potential reason for the difference in group performance is that the "musical" participants in group 2 might have made more effort post-implantation to rehabilitate themselves, which is supported by their listening habits at T1. This could explain why group 1 appear to show more improvement on the outcome measures following IMAP use than for group 2 (Figure 5). However, potential differences between the groups in terms of their pitch acuity and cognitive abilities cannot be ruled out. Future trials could allocate participants to groups where the randomization is weighted based on such characteristics to ensure the groups are more evenly matched (see for example, Treasure & MacRae, 1998).

The difference in improvement between the groups could also be partly due to a combination of a labeling effect and differences in how long each of the groups had to wait to use the IMAP. Prior to T1, participants received an information sheet, which explained that they would be randomly allocated to "group 1" and would receive the IMAP at T1, or "group

2" the "control group" and would not be given the IMAP until after a period of 12 weeks, at which point the groups would swap over. The label "control group", which was only applied to group 2, could have biased this groups' test performance and possibly also their attitude towards the IMAP. In addition, this groups' motivation for participating in the trial might have waned over the 12-week period in which that they had to wait for the IMAP. In designing this study, a blind, placebo controlled trial was considered to ensure equal motivation between the groups, as in the study of Moreno et al. (2009), who explored the influence of musical training on the linguistic abilities of children using painting training as a control. However, due to the participatory design approach taken in developing the IMAP and general public knowledge of the research, it was difficult to conceive of a convincing and yet ineffectual placebo for group 2 for a clean comparison, and impractical to blind participants as to which group they were in. In addition, it is likely that the adult participants in this study would have questioned the relationship between the music-related outcome measures and any non-aural training.

Both groups are also unlikely to be representative of the wider population by the very nature that they volunteered to take part in the trial of the IMAP and are therefore more likely to have an interest in music and higher motivation to improve their music perception abilities.

Conclusions and Summary

The main observations of this study can be summarized as follows:

- Following training with the prototype IMAP, both groups showed improved instrument recognition, consistent with the findings of Fujita and Ito (1999), Gfeller, Witt, Adamek, et al. (2002), Gfeller, Witt, Woodworth, et al. (2002), Petersen et al. (2012), and Driscoll (2012) suggesting that with practice, adult CI users can relearn the timbre of musical instruments.

- Retention of learning for instrument recognition (measured for group 1 only) was observed 12 weeks following training with the IMAP.

- Change in MCI performance from T1 to T2 between the groups approached statistical significance. However, scores for group 2 might have been limited by ceiling effects.

- Post-training, a statistically significant improvement in UK-matrix SRT was observed for group 1, but not for group 2, potentially reflecting prior differences in group characteristics that were not measured.

- Post-training sound quality ratings for group 1 increased whereas the ratings for group 2 did not, potentially reflecting differences in music exposure between the groups prior to training.

- A general, but non-statistically significant increase in music listening habits was observed for both groups throughout the course of the trial. Feedback suggests that the IMAP had a positive impact on participants' lives not captured by the outcome measures.

- Overall adherence was good with participants spending on average 41 minutes per session, and 14 out of the 16 participants who attended all three music perception assessment appointments completing ≥20 sessions.

Following this study, the prototype IMAP has been further refined based on qualitative analysis of the feedback and it is now freely available online[4] so that it can easily be maintained and updated. The next stages of this research are to further evaluate and document the efficacy and effectiveness of the IMAP so that ultimately, patients and professionals can make an informed decision as to how much time they invest in undertaking music aural rehabilitation.

---

[4] The Interactive Music Awareness Program (IMAP) is available online at www.MoreFromMusic.org

Author Note

Correspondence concerning this article should be addressed to Rachel M. van Besouw Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. E-mail: rvb@isvr.soton.ac.uk

References

BENCH, J., KOWAL, Â., & BAMFORD, J. (1979). The Bkb (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13*, 108-112.

BRAND, T., & KOLLMEIER, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimate for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America, 111*, 2801-2810.

DE BOER, J., & THORNTON, A.R.D. (2008). Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *Journal of Neuroscience, 28*, 4929-4937.

DRISCOLL, V.D. (2012). The effects of training on recognition of musical instruments by adults with cochlear implants. *Seminars in Hearing, 33*, 410-418.

DUNN, G., MARACY, M., DOWRICK, C., AYUSO-MATEOS, J.L., DALGARD, O.S., PAGE, H., LEHTINEN, V., CASEY, P., WILKINSON, C., VÁZQUEZ-BARQUERO, J.L., & WILKINSON, G. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry, 183*, 323-331.

EYSENBACH, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7, e11.*

FUJITA, S., & ITO, J. (1999). Ability of Nucleus cochlear implantees to recognize music. *Annals of Otology, Rhinology, and Laryngology, 108*, 634-640.

GALVIN III, J.J., FU, Q-.J., & NOGAKI, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear and Hearing, 28*, 302-319.

GENESIS. (2009). History and description of loudness models: loudness toolbox for Matlab. *Genesis S.A. document IB/RP/10003*. Retrieved from: www.genesis.fr

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, 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, 120-134.

GFELLER, K., OLSZEWSKI, C., RYCHENER, M., SENA, K., KNUTSON, J.F., WITT, S., & MACPHERSON, B. (2005). Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear and Hearing,* 26, 237-250.

GFELLER, K., TURNER, C., OLESON, J., ZHANG, X., GANTZ, B., FROMAN, R., & OLSZEWSKI, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing,* 28, 412-423.

GFELLER, K., WITT, S., ADAMEK, M., MEHR, M., ROGERS, J., STORDAHL, J. & RINGGENBERG, S. (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology,* 13, 132-145.

GFELLER, K., WITT, S., STORDAHL, J., MEHR, M., & WOODWORTH, G. (2000). The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *Journal of the Academy of Rehabilitative Audiology,* 33, 115-138.

GFELLER, K., WITT, S., WOODWORTH, G., MEHR, M., & KNUTSON, J. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology, and Laryngology,* 111, 349-356.

GLASBERG, B.R., & MOORE, B.C.J. (2002). Model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society,* 50, 331-341.

HAGERMAN, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology,* 24, 71-77.

HAGERMAN, B., & KINNEFORS, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scandinavian Audiology,* 24, 71-77.

LASSALETTA, L., CASTRO, A., BASTARRICA, M., PÉREZ-MORA, R., HERRÁN, B., SANZ, L., DE SARRIÁ, M.J., & GAVILÁN, J. (2008). Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringológica Española,* 59, 228-234.

LEAL, M.C., SHIN, Y.J., LABORDE, M.-L., CALMELS, M.-N., VERGES, S., LUGARDON, S., ANDRIEU, S., DEGUINE, O., & FRAYSSE, B. (2003). Music perception in adult cochlear implant recipients. *Acta Oto-laryngologica,* 123, 826-835.

LEVITT, S.D., & LIST, J.A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal Applied Economics,* 3, 224-238.

LOOI, V. (2008). The effect of cochlear implantation on music perception. *Otorinolaringologia,* 58, 169-190.

LOOI, V., GFELLER, K., & DRISCOLL, V. (2012). Music appreciation and training for cochlear implant recipients: a review. *Seminars in Hearing,* 33, 307-334.

LOOI, V., & SHE, J. (2010). Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *International Journal of Audiology,* 49, 116-128.

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, 503-518.

MCARDLE, R.A., WILSON, R.H., & BURKS, C.A. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *Journal of the American Academy of Audiology,* 16, 726-739.

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

MIRZA, S., DOUGLAS, S.A., LINDSEY, P., HILDRETH, T., & HAWTHRONE, M. (2003). Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear Implants International,* 4, 85-95.

MORENO, S., MARQUES, C., SANTOS, A., SANTOS, M., CASTRO, S.L., & BESSON, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex,* 19, 712-723.

OLIVER, B.R., VAN BESOUW, R.M., & NICHOLLS, D.R. (2012). The 'Interactive Music Awareness Program' (IMAP) for cochlear implant users. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 248-250.

PARBERY-CLARK, A., SKOE, E., LAM, C., & KRAUS, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing,* 30, 653-661.

PERETZ, I., GAUDREAU, D., & BONNEL, A-.M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition,* 26, 884-902.

PETERSEN, B., MORTENSEN, M.V., HANSEN, M., & VUUST, P. (2012). Singing in the key of life: a study on effects of musical ear training after cochlear implantation. *Psychomusicology: Music, Mind, and Brain,* 22, 134-151.

PHILIPS, B., VINCK, B., DE VEL, E., MAES, L., D'HAENENS, W., KEPPLER, H., & DHOOGE, I. (2012). Characteristics and determinants of music appreciation in adult CI users. *European Archives of Otorhinolaryngology,* 269, 813-821.

RUSSO, N.M., NICOL, T.G., ZECKER, S.G., HAYES, E.A., & KRAUS, N. (2005). Auditory

training improves neural timing in the human brainstem. *Behavioural Brain Research,* 156, 95-103.

STACEY, P.C., RAINE, C.H., O'DONOGHUE, G.M., TAPPER, L., TWOMEY, T., & SUMMERFIELD,

A.Q. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology,* 49, 347-356.

TREASURE, T., & MACRAE, K.D. (1998). Minimisation: the platinum standard for trials? *The BMJ,* 317, 362-363.

TYLER, R.S., PARKINSON, A.J., WOODWORTH, G.G., LOWER, M.W., & GANTZ, B.J. (1997).

Performance over time of adult patients using the Ineraid or Nucleus cochlear implant. *Journal of the Acoustical Society of America,* 102, 508-522.

VAN BESOUW, R.M., NICHOLLS, D.R., OLIVER, B.R., HODKINSON, S.M. & GRASMEDER, M.L.

(2014). Aural rehabilitation through music workshops for cochlear implant users. *Journal of the American Academy of Audiology*, 25, 311-323.

WAGENER, K., JOSVASSEN, J.L., & ARDENKJÆR, R. (2003). Design, optimization and

evaluation of a Danish sentence test in noise. *International Journal of Audiology,* 42, 10-17.

TABLE 1

Participant Characteristics

ID	Group	Age (yrs)	Sex	Pre/post lingual deafness	Listening mode	Duration post i.t. (mo)	BKB in quiet (%)	Prior music training
4	1	73	M	post	U	18	>60	none
5	1	40	F	post	U	18	81	none
7	1	66	M	post	U	94	92	none
11	1	68	M	post	U	41	94	none
12	1	77	M	post	U	85	93	none
14	1	52	M	pre	Bm	20	68	none
15	1	51	F	post	U	80	94	none
17	1	42	F	post	Bl	77	99	none
19	1	55	F	post	U	93	100	none
1	2	46	F	post	U	51	100	none
6	2	67	F	post	U	68	84	none
8	2	68	F	post	U	45	98	<5 yrs
9	2	59	M	post	Bm	33	100	<5 yrs
10	2	51	F	uncertain	U	23	60	>5 yrs
13	2	60	F	post	Bm	47	99	<5 yrs
18	2	65	F	post	Bm	61	100	<5 yrs

Note. Group 1 used the IMAP between T1 and T2, and group 2 used the IMAP between T2

and T3. Duration post i.t.: duration in months since the initial tuning of the first implant. BKB

in quiet: most recent score for the BKB sentence test in quiet; the score listed as ">60"

indicates that the performance of the individual is known to be greater than 60%, but an

actual BKB score in quiet was unavailable as the individual is now tested using the adaptive

BKB speech-in-noise test. Listening mode: U = unilateral, Bm = bimodal, Bl = bilateral.

Prior music training: number of years of formal music training or study prior to implantation.

TABLE 2

Music Sound Quality Stimuli

Order	Genre	Artist	Track title	Album	Release	Label
1	Blues	Savoy Brown	You're In For A Big Surprise	Bring It Home	1994	Viceroy
2	Classical	Haydn Trio Eisenstadt	Keyboard Trio No. 26 in F sharp minor, Hob.XV:26: III. Tempo di minuet	Haydn: Trios for Piano, Violin and Cello	2010	Capriccio
3	Country	Matraca Berg	I Must Have Been Crazy	Lying to the Moon	1990	RCA
4	Jazz	Thelonious Monk	Straight, No Chaser	Straight, No Chaser	1996	Columbia /Legacy
5	Pop	Candy Butchers	What I Won't Give	Making Up Time	2006	Good Morning Monkey Records

FIGURE 1   Randomized crossover design. Group 1 received the IMAP first, followed by a 12-week retention of learning phase. Group 2 were given no intervention for the first 12 weeks and then were given the IMAP to use. Both groups were required to attend three 2-hour music perception assessment appointments at T1, T2 and T3.

FIGURE 2   UK-matrix sentence test response options. Using a touch screen, participants were required to select a name, verb, numeral, adjective and object from 10 alternatives (e.g. selected options shaded for "Thomas has three green beds").

FIGURE 3   Response options in the 9-alternative forced choice MCI test.

FIGURE 4   Instrument recognition test response options.

FIGURE 5   Music perception measures and sound quality ratings for group 1 who used the IMAP between T1 and T2 and for group 2 who used the IMAP between T2 and T3. Panels: (a) UK-matrix SRT, (b) MCI performance, (c) instrument recognition, and (d) sound quality ratings. White boxes: pre-IMAP performance, shaded boxes: post-IMAP performance, box: interquartile range, upper and lower whiskers: top and bottom 25% of scores excluding outliers, ○: outlier between 1.5 and 3 times the interquartile range, *: statistically significant change in performance/ratings between sessions.

FIGURE 6   Responses to the question "how often do you choose to listen to music?" at

T1, T2 and T3 for groups 1 and 2. 'Less' means 'less than monthly'. At T2

and T3 participants were asked how often they had chosen to listen to music

since their last appointment. If a participant had just completed the training

phase, they were asked to consider how often they had chosen to listen to

music in addition to the music that they had been exposed to in the IMAP.

FIGURE 7   Bars: total time spent using the IMAP by each participant. Dashed line: total

time spent using the IMAP assuming a session length of 30 minutes. Solid

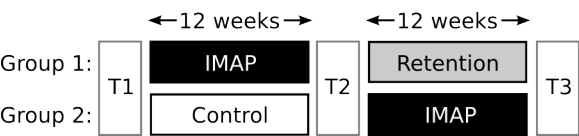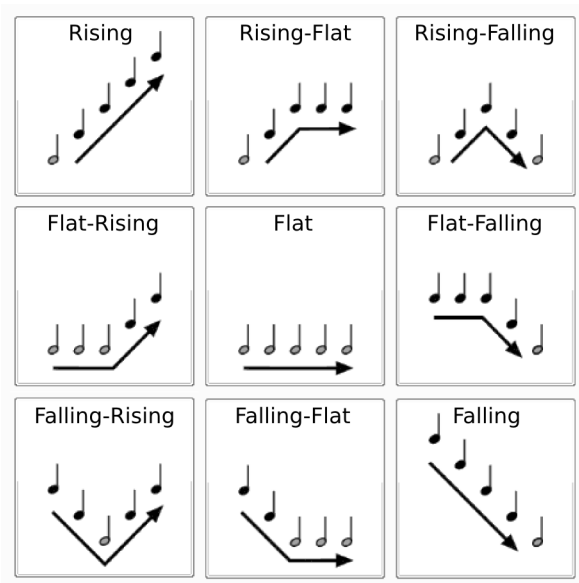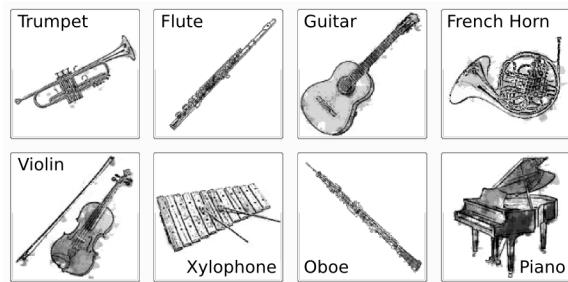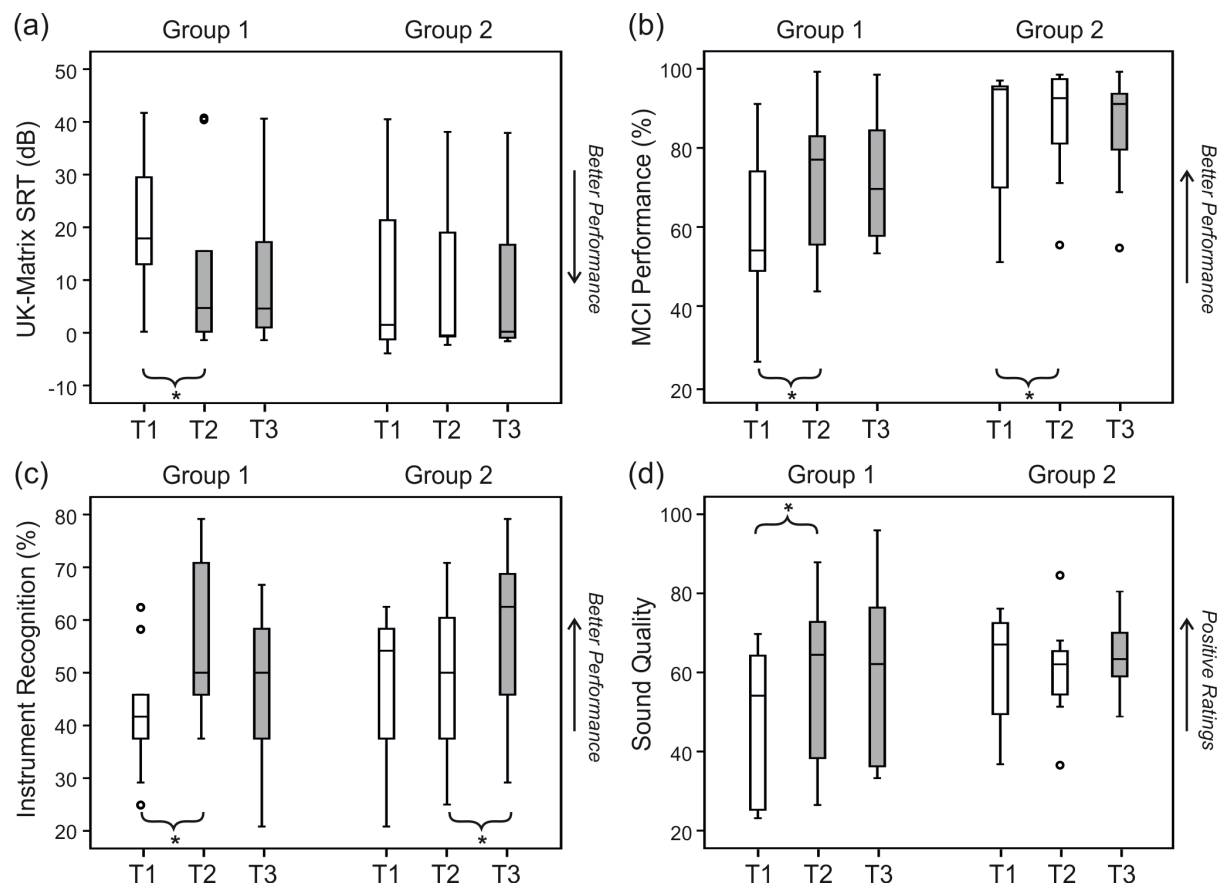line: number of IMAP sessions completed.

FIGURE 1

FIGURE 2

Peter	got	three	large	desks
Kathy	sees	nine	small	chairs
Lucy	bought	five	old	shoes
Alan	gives	eight	dark	toys
Rachel	sold	four	thin	spoons
Barry	likes	six	green	mugs
Steven	has	two	cheap	ships
Thomas	kept	ten	pink	rings
Hannah	wins	twelve	red	tins
Nina	wants	some	big	beds

FIGURE 3

FIGURE 4

FIGURE 5

FIGURE 6

FIGURE 7