Full-Length Article

Multidisciplinary Perspectives on Music Perception and Cognition for Cochlear Implant Users

Alexander Chern1,2, Iliza M. Butera3

1 Department of Otolaryngology - Head and Neck Surgery, Columbia University Irving Medical Center
2 Department of Otolaryngology - Head and Neck Surgery, Weill Cornell Medicine
3 Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN

Abstract

For over 30 years, cochlear implants (CIs) have been successfully providing sound and speech perception to individuals who suffer from severe-to-profound sensorineural hearing loss. Despite many recent advances in CI technology, significant challenges remain for users, including speech perception in noisy environments, identifying vocal emotion, and perhaps most notably, music perception and appreciation. Moreover, pediatric cochlear implant users often demonstrate a slower and more variable language development trajectory compared to their normal hearing peers, which is in part due to the imperfect hearing restoration by these devices. In this brief report, we discuss multidisciplinary perspectives on music perception and cognition for CI users, as well as how they can be employed to improve the cochlear implant listening experience. We divide these strategies into two categories—a top-down approach (e.g., employing therapeutic measures to help train the CI user’s brain to fully reap the benefits of cochlear implantation) and a bottom-up approach (e.g., improving the auditory input through developing new technology, creating individualized programming strategies, and developing music specifically tailored for CI users). These individualized, yet multidirectional approaches will help create a functionally-integrated system that supports robust processing of complex sounds, which is essential for many everyday tasks.

Keywords: cochlear implant, hearing loss, music cognition, music perception

Introduction

Cochlear implants (CIs) are a surgical intervention for severe-to-profound sensorineural hearing loss, and with over half a million users worldwide, they are by far the world’s most successful neuroprosthetics. Over the past few decades, advances in CI processing systems have yielded substantial gains in speech perception such that most CI users are able to achieve satisfactory—or even near perfect—speech recognition scores in quiet environments.[1] However, despite considerable technological and surgical advances in the field, significant limitations remain for CI users, including speech understanding in noisy environments, detection of vocal emotion, and notably, music perception and enjoyment. Such tasks require processing of spectrally-complex sounds that are ineffectively conveyed by CIs. Moreover, children with cochlear implants often demonstrate a slower and more variable language development trajectory compared to their normal-hearing peers.2 This may be due to several reasons, including spectrally-impoveryed acoustic signals provided by CIs, as well as lack of access to auditory speech signals for prelingually deafened children until their CIs are activated.

In this review, we highlight technological limitations of CIs, the challenges that users face, and recent research that may improve the CI experience via multidisciplinary approaches spanning the fields of music composition/production, biomedical engineering, otolaryngology, audiology, and neuroscience (Fig. 1). These potential advancements and therapies are broadly organized from “bottom-up” to “top-down” strategies, meaning they can improve either the quality of ascending auditory information or the higher cognitive processes that shape perception.
Figure 1. Multidisciplinary approaches to improving music perception and cognition with a cochlear implant. We organize possible interventions from external factors to those involving the inner ear and finally, higher cognitive processes. In other words, from “bottom-up” to “top-down” approaches.

How hearing works

The auditory pathway is composed of two parts, the peripheral auditory system and the central auditory system. The peripheral auditory system is made up of several components. The outer ear is composed of the auricle (or pinna) and external auditory canal, which serve to funnel and focus sound waves onto the tympanic membrane (or eardrum). The tympanic membrane, which separates the outer and middle ear, then vibrates, transmitting the acoustic energy from sound waves to the three tiny bones of the middle ear - the ossicles. These three ossicles (malleus, incus, and stapes) then further transmit this vibration, thus efficiently transferring sound waves from air to a fluid medium (perilymph) in the inner ear through an area advantage (i.e., hydraulic amplification) and a lever effect. The cochlea of the inner ear receives this mechanical signal through vibratory pressure from the stapes footplate applied to the oval window, which creates a traveling fluid wave that causes hair cell deflection along the base to the apex of the cochlea, stimulating bipolar neurons of the spiral ganglion that form the auditory nerve (part of the eighth cranial nerve). Thus, the inner ear helps convert this mechanical signal into an electrical signal that can be carried by the auditory nerve through the central auditory system to the brain. The central auditory system begins at the cochlear nuclei in the brainstem, which receive spiral ganglion axons of the spiral ganglion that form the auditory nerve (part of the eighth cranial nerve). Thus, the inner ear helps convert this mechanical signal into an electrical signal that can be carried by the auditory nerve through the central auditory system to the brain. The central auditory system begins at the cochlear nuclei in the brainstem, which receive spiral ganglion axons of the spiral ganglion that form the auditory nerve (part of the eighth cranial nerve). Thus, the inner ear helps convert this mechanical signal into an electrical signal that can be carried by the auditory nerve through the central auditory system to the brain.

How cochlear implants work

Though CIs and hearing aids (HAs) are both medical devices used to treat hearing loss, they are very different. HAs amplify (i.e., increase the volume of) sounds and send them to the brain through the entire auditory pathway, as described above. HAs may benefit individuals with a wide range of mild to severe hearing loss. On the other hand, CIs bypass damaged portions of the ear and directly stimulate auditory spiral ganglion cells (Fig. 2B); CIs are typically used by individuals with severe-to-profound sensorineural hearing loss who derive minimal benefit from HAs. A CI consists of several components (Fig. 2A). Sound from the environment is detected by a microphone near the opening of the ear canal. The adjacent processor translates the frequency and amplitude of sound waves into electrical pulses. These digital signals are sent by the transmitter to an internal receiver/stimulator implanted under the skin. Lastly, the electrical pulses are delivered to the inner ear where auditory spiral ganglion cells are excited in a tonotopic manner (i.e., from high to low frequencies spiraling inward from the base to the apex of the cochlea) via the electrode array (Fig. 2B). This entire process takes place on the scale of about 10 ms and establishes a rudimentary sense of hearing.
Why music perception is difficult with an implant

Processing an acoustic signal (e.g., speech or music) involves both temporal and spectral cues. Temporal envelope processing is the ability to resolve signal changes over time; it helps us distinguish words, speech rhythm and stress, as well as musical rhythm. Spectral information is the fine structure of the frequency composition of sounds. Therefore, spectrotemporal processing is the ability to resolve component frequencies over time. This is analogous to how a Fourier transform is used to resolve a time-domain function into a frequency spectrum. While spectral resolution helps us distinguish some aspects of speech (e.g. voice quality and intonation), words themselves can be conveyed quite well via temporal cues. In contrast, crucial aspects of music like pitch and timbre are encoded by spectral cues.

The discrete number of electrodes in a CI limits spectral resolution in part due to the excessive spread of intracochlear electric current around each electrode contact. This phenomenon of “channel interaction” causes interference between channels and the inability to resolve more than 8-10 independent channels of information.[5–7] See Figure 3 for an example of how cochlear implant processing degrades a melody. We also provide a supplementary video illustrating the phenomenon.

Figure 3. An example of how cochlear implant processing degrades a melody. (A) We selected a short classical melody for its simple structure and recognizability in its original form, which we demonstrate using a sampled piano sound via GarageBand software. (B) Though CIs accurately convey the overall waveform of music via high temporal resolution, the harmonics, timbre, and other crucial spectral components are absent (C), which drastically affects how this melody sounds to users. For a better view of the acoustic spectrogram (D), a segment (C, black box) is expanded to show the visual patterns of stacked lines that represent overtones (including harmonics) and define the timbre of an instrument. In the expanded view of the CI simulation spectrogram (E), these visual patterns are largely absent. To fully appreciate the importance of this missing spectral information, we recommend downloading the supplementary video to take a listen.

Click **HERE** for video

https://drive.google.com/open?id=1jAJ_LTmP-Qk3qd72ISR24wie2Yb2z4aV

**Bottom-up approaches to improving music perception for CI users**

**Tailored music composition and production**

One possible intervention is to better design music for the known limitations of cochlear implant processing. For instance, melody is an important aspect of music; however, standard melodies typically consist of neighboring notes with such small differences in pitch that several may be filtered to the same electrode for a CI user. This means that very obvious differences between notes for acoustic listeners—such as A7 to B7 (i.e., 3,520 Hz to 3,951 Hz)—may be filtered to the same electrode within the cochlea (e.g., electrode 14, Fig. 3C inset). As a result, these two notes will sound identical to CI users and will be less melodic. Although the minimum perceptible difference in pitch varies between CI users, a difference of 3 to 8 semitones is typically required for discrimination.[9–11] This range is much larger than normal hearing listeners’ ability to resolve differences of less than one semitone. Furthermore, polyphonic melodies with overlapping notes are far too complex and cacophonous for CI users. Instead, utilizing simple, monophonic melodies (i.e., with only individual notes in sequence) can sound more melodic with electric hearing via CIs.

This limitation in spectral resolution affects melody perception for CI users as well as timbre discrimination. Timbre describes the difference in instrument sound qualities when, say, a piano and violin play the same note at the same volume and for the same duration. Although the sources of these overlapping notes can be clearly distinguished by typical listeners, CI users may only perceive a single, ambiguous sound. In fact, timbre discrimination can be so poor that it is possible for even a flute and a snare drum to sound the same. In contrast, simple synthesized sounds such as square waves that contain regular harmonics and no reverb may be more preferable for CI users than the complex, yet indistinguishable timbres of acoustic instruments like pianos and violins.

Rhythm and lyrics are much more accessible musical components for CI users. Similar to speech, they rely more on temporal information that implants can better encode. Therefore, music that emphasizes rhythm and lyrics while simplifying timbre and melodies is more likely to sound “musical” to CI users.[12] Additionally, creating short loops within a song can better synchronize neural firing in the auditory brainstem and enhance musicality, even for typical listeners. This so-called “sound-to-song” effect can be illustrated by listening to a short loop (~1 s long) of a sound, which becomes increasingly more musical over time.[13] In the same way, looping otherwise difficult-to-perceive musical
motifs could be perceptually enhanced for CI users’ listening enjoyment.

**Intraoperative, surgical considerations**

Surgical considerations can also be made to optimize spectral fine structure processing. Typically, a CI is blindly inserted into the scala tympani (one of the bony cavities of the inner ear) through either the round window (a membrane that separates the middle ear from the inner ear) or through a cochleostomy (the creation of a hole adjacent to the round window in the cochlea, as seen in Figure 4).

In recent years, there has been an increasing emphasis on minimizing intracochlear trauma and optimizing electrode contacts closer to spiral ganglion cell targets. These surgical factors are known to affect post-implantation outcomes, so it is important to establish techniques and strategies that provide real-time, intraoperative feedback regarding CI insertion and positioning. For example, recent studies have investigated the utility of real-time, intraoperative electrocochleography (ECochG) in providing information regarding intracochlear trauma[14] and the location of the electrode array within the cochlea.[15,16] Proper insertion of the electrode array into the scala tympani (i.e., without crossover into scala vestibuli; see Figure 2D and 2E) has been associated with better hearing outcomes and decreased intracochlear trauma, which may facilitate residual hearing preservation.[17] Use of a flexible, shorter electrode array may also cause less harm to the healthy portion of the cochlea in individuals with residual low frequency hearing because this segment is farthest away from the round window (or cochleostomy) insertion site. Studies have shown a correlation of both intracochlear damage[18] and worse hearing preservation[19] with increased depth of insertion.

These strategies help preserve residual low frequency hearing, which enable patients to take advantage of electroacoustic stimulation (EAS), where high-frequency information is provided by a CI and low frequencies can be delivered with acoustic amplification with a HA in the same ear.[20] EAS has been shown to be beneficial for tasks that require a relatively high spectral resolution, including speech perception in noise[21] and pitch perception.[22]

**Postoperative, audiological programming**

After CI surgery, patients see their audiologist who activates and programs the device. This individualized program involves several parameters including frequency allocations and stimulation levels. Programming optimizes the CI user’s listening experience by fitting acoustic input into the reduced dynamic range of electric hearing through an implant. Electrode arrays are designed so each electrode stimulates nerve pathways corresponding to a spectral bandwidth of frequencies. However, a one-size-fits-all approach results in decreased programming efficiency, since there are patient-to-patient variations in final electrode placement. High resolution CT imaging can help characterize electrode positions in vivo, and thus facilitate an individualized image-guided mapping strategy that has been shown to enhance spectral fine structure processing via selective deactivation of electrodes to reduce channel interaction[7] and also improve pitch perception outcomes by reducing place-pitch mismatch.[24] Another precision medicine strategy being investigated is the Fitting to Outcomes eXpert (FOX) trial; this investigates the utility of FOX, an artificial intelligence software tool, in individualized CI programming for each patient based on specific outcome measures.[25] Such tools will enable audiologists to better customize CI programming parameters to optimize CI pitch discrimination.

**Top-down methods for improving music perception for CI users**

**Musical training-induced benefits in music perception**

Since most CI users today can obtain good speech perception in quiet, there has been a shift towards more complex auditory rehabilitation paradigms, including musical training. For example, benefits in melodic contour identification, as well as pitch and timbre perception have been observed with longitudinal training interventions. [26] This type of intervention can be implemented to potentially improve music perception and appreciation in CI users, regardless of their musical abilities, type of device, or experience with the CI.[26]

**Rhythmic priming effect on language**

Children with cochlear implants often demonstrate a slower and more variable language development trajectory compared to their normal peers;[2] this may be due to several reasons,
including phonological impairments as a result of the spectrally-impo
erived acoustic signals provided by CIs (e.g., impaired identification of short grammat
cal morphemes), as well as lack of access to auditory speech signals for prelingually deafened children until their CIs are activated.

A large body of literature has explored shared neural resources for music and language, notably in rhythm/timing and grammar (i.e., syntactic) processing. Behavioral studies conducted in children have also demonstrated the positive influence of rhythmically regular musical stimulation on grammar task performance in children with normal development and developmental language delay. Bedoin et al. showed that this observed improved language processing state (i.e., short-term benefits in language processing induced by rhythmically regular musical stimulation in the laboratory) can be extended to a permanent trait (i.e., long-term effect) when incorporated into language training/rehabilitation programs in pediatric CI users, a population known to have language deficits. Thus, regular musical stimuli may enhance grammar processing when incorporated into language therapy in pediatric CI users.

Multisensory enhancements for music listening

Degraded auditory input is common to all CI processors and, like hearing loss, prompts attentional focusing on other sensory modalities. Fortunately, speech is typically an audiovisual experience wherein coincident orofacial articulations can considerably boost intelligibility over auditory-alone listening. This is also true for typical listeners who can benefit from visual speech cues to communicate in otherwise unintelligible auditory signal-to-noise ratios. For many CI users, communication by phone is only possible through audiovisual formats such as FaceTime. When faced with impaired auditory inputs for speech as well as music, the incorporation of visual cues is an intuitive and effective compensatory strategy. Practically, this can mean live music may be more enjoyable for CI users than recorded music that lacks any visual cues from the musicians’ gestures. Furthermore, designing concurrent visual cues for music listening via new software designs could also improve listening enjoyment for CI users. In short, paired visuals and/or haptics with auditory listening could prove useful for facilitating the natural enhancement that multisensory listening affords.

Discussion and future directions

Historically, CI research has been focused on optimizing CI processing systems, types of electrode arrays, and surgical approaches for implant placement. Complex auditory processing (e.g., music perception and speech in noise) remains a challenge for patients, and will continue to be addressed through a variety of approaches such as the external input/translation of acoustic sounds, peripheral encoding within the cochlea, as well as more central processing interventions. Such strategies require collaboration across disciplines including (A) music composition/production, biomedical engineering, (B) otolaryngology, audiology, and (C) neuroscience (Figure 1). Given a broad trend towards precision medicine (e.g., targeted cancer therapy), taking an individualized approach in optimizing the CI experience (e.g., CT-guided CI mapping, tailoring music composition to a CI user) may prove beneficial for improving spectral fine structure processing and possibly music perception and appreciation. Moreover, shared neural pathways for music and language processing may mean that improvements in music perception may generalize to also improving the CI linguistic experience.

Considering that we all live and work in complex listening environments, improving speech-in-noise perception has many practical benefits. For example, in a busy clinical environment like in the intensive care unit or in the operating room, our brains are saturated with noises and other complex sounds such as alarms with varying degrees of urgency. For CI users, communicating in this environment may be particularly challenging due to the low spectral resolution of CI-mediated sounds. As a result, we must be conscious of how complex sounds are frequently present in day-to-day life, yet may be inaccessible for CI users.

It should be noted that while the strategies discussed in this article may benefit music perception and appreciation, not all have been validated through objective measures that assess musical (e.g., pitch or timbre) perception. Clearly, musical perception is required for music appreciation, and it is important to also establish whether these strategies have further clinical significance to everyday listening environments.

Conclusion

Interdisciplinary, individualized approaches are necessary for improving music listening for cochlear implant users. Collectively, we should explore external factors of music composition and production as well as implant design and processing, and intraoperative surgical considerations for optimizing placement of the electrode array within the cochlea. Finally, as the neural representation of music is centrally processed by the brain, there are potential strategies in integrating auditory information with other sensorimotor information (e.g. visuals and haptics), as well as training the brain through auditory rehabilitation to more fully benefit from CI-mediated hearing.

Acknowledgements
We thank Bob Dwyer and Jourdan Holder for helpful discussions of many topics in this manuscript, and in particular for feedback regarding the depictions in Fig. 2 and the CI frequency mappings illustrated in Fig. 3c. This work was supported in part by the National Institutes of Deafness and Communication Disorders award No. 5P31DC015956 (Butera) in addition to NIH funding from R01DC008408 and R01DC014462 to support the database from which CT images in Fig. 2CE were accessed. The contents are solely the responsibility of the authors and do not necessarily represent the official views of any of these funding sources.

References


Biographical Statements
Alexander Chern, MD is an Otolaryngology—Head & Neck Surgery resident physician at NewYork-Presbyterian Hospital (Columbia University Irving Medical Center/Weill Cornell Medicine).

Iliza Butera received her PhD in Neuroscience from Vanderbilt University, and is an analyst at Artiphon.