COCHLEAR IMPLANTS CAN TALK BUT CANNOT SING IN TUNE

Jeremy Marozeau, Ninia Simon and Hamish Innes-Brown
The Bionics Institute, East Melbourne, Australia
jmarozeau@bionicsinstitute.org

The cochlear implant is rightfully considered as one of the greatest success stories in Australian biomedical research and development. It provides sound sensation to hundreds of thousands of people around the world, many of whom are able to understand and produce speech. The device was developed in order to optimize speech perception, and parameters such as the choice of frequency bands and signal processing used were chosen in order to maximise perceptual differences between speech vowels. However, these settings are far from being suited for the perception of music, which might partly explain why many cochlear implant recipients cannot enjoy music through their implant.

INTRODUCTION

The cochlear implant was developed simultaneously in the sixties in France, the US and Australia [1]. Initially the cochlear implant , CI, was designed to increase the ability of the profoundly deaf to navigate through the environment by providing them with sound sensation. Due to the complexity of the cochlea it was thought that improvements in speech perception or directional hearing for future CI recipients would be of minor impact [2]. However, speech understanding outcomes improved rapidly and nowadays many CI users are able to understand words in sentences in quiet listening environments without any other aid [2, 3]. Understanding speech is important for verbal communication; however, many other sounds are also important for non-verbal communication and understanding of the auditory environment. Music is an obvious example. Listening to music forms important social bonds, and is ubiquitous in our society, from the movies to the supermarket. Music is invariably present in situations where group and community reinforcement occurs (e.g. concerts, weddings, funerals, protests, State occasions). A survey of the musical habits of over 100 CI listeners found that music was generally less enjoyable post-implantation [4]. Although there is a wide range of musical abilities among CI users, many display major limitations in their ability to perceive and enjoy music [for a review see 3, 5, 6].We will argue in this paper that the strong focus on speech processing has been detrimental to the perception of music.

THE COCHLEA

The cochlea is a spiral structure with a shape similar to the Nautilus shells sometimes found washed up on beaches, and its job is to convert mechanical vibrations within the cochlea into electrical pulses in the auditory nerve. The cochlea is embedded in the temporal bone, and in a healthy ear contains rows of hair cells lined up along its length. These hair cells stimulate auditory nerves when they are moved by vibrations in the basilar membrane, in which they are mounted. The basilar membrane has mechanical properties causing it to resonate at different frequencies along its length. The hair cells are thus set in motion at different points along the membrane depending on the frequency of the sound. If hair cells close to the middle ear vibrate, a high-pitched sound is heard. The pitch gradually gets lower as the region of hair cells that vibrate get deeper into the cochlear.

Table 1. Left panel: Electrode position (the distance from the apex in mm) for a patient with 33 mm cochlear length and an insertion of 20 mm, the corresponding frequency according to the Greenwood function. Right panel: Frequency allocations of a standard Cochlear© cochlear implant.

Distance From Apex [mm]	Frequency Predicted by Greenwood function. [Hz]	Electrode Number	Lower Freq [Hz]	Upper Freq [Hz]
13	966	22	188	313
13.9	1124	21	313	438
14.8	1305	20	438	563
15.7	1512	19	563	688
16.6	1748	18	688	813
17.5	2017	17	813	938
18.5	2326	16	938	1063
19.4	2678	15	1063	1188
20.3	3080	14	1188	1313
21.2	3540	13	1313	1563
22.1	4065	12	1563	1813
23.0	4665	11	1813	2063
23.9	5350	10	2063	2313
24.8	6133	9	2313	2688
25.7	7028	8	2688	3063
26.6	8050	7	3063	3563
27.5	9218	6	3563	4063
28.5	10552	5	4063	4688
29.4	12076	4	4688	5313
30.3	13818	3	5313	6063
31.2	15807	2	6063	6938
32.1	18080	1	6938	7938

Based on the work of the Nobel laureate Georg von Békésy, Greenwood [7] developed a function that predicts the position of maximum excitation on the cochlea as a function of the frequency of the input sound. Table 1 shows some of these relationships for a patient with a 33 mm long cochlea. For example, when a patient with a CI is presented with a 700 Hz tone, electrode 18 will be activated. Stimulation at this electrode will produce a maximum gradient potential around 16.6 mm from the apex. In normally-hearing listeners, this region is activated by a tone with a frequency of 1748 Hz.

There are around 3500 rows of hair cells along the length of the basilar membrane. If they are damaged, the auditory neurons cannot be excited, leading to a loss of the sensation of hearing.

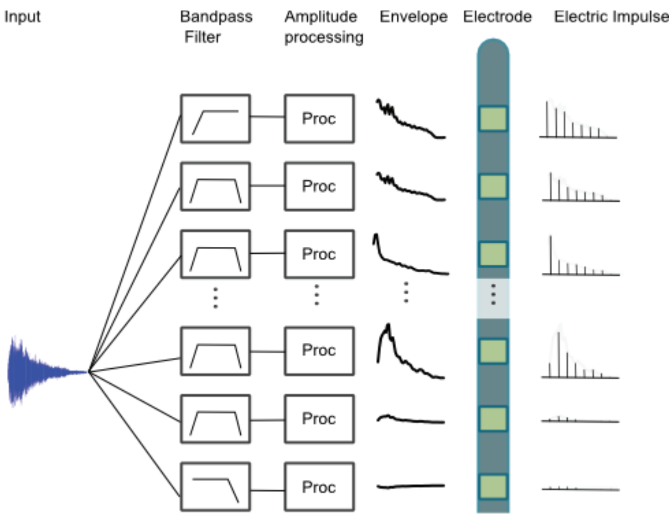

Figure 1. Simplified block diagram of a cochlear implant. First the sound is decomposed into frequency bands. Then the output of each band is processed through compression and an envelope detector. A series of electrical impulses at fixed rate modulated with the envelope will be produced on selected electrodes. The frequency allocation of each electrode is predetermined by the frequency of each channel (represented in Table 1 for Cochlear Device.)

THE COCHLEAR IMPLANT

A cochlear implant largely replaces the function of the outer, middle, and most of the inner ear – up to the level of the auditory nerve. It consists of two main parts. First, the sound processor is worn externally, and hooks behind the ear. It contains microphones, batteries, and a miniaturized computer system that converts the acoustic signal received at the microphones into a series of electric pulses according to a programmable software algorithm called a 'strategy.' Second, the implant itself is implanted in the mastoid bone behind the ear. It receives power, as well as the electrical signals from the sound processor via a wireless link through the skin. At the end of the implant is a very fine linear array of up to 22 electrodes, which is inserted about half-way into the spiral-shaped cochlea. These electrodes stimulate the auditory nerve, thus replacing the function of the hair cells that are lost or damaged in sensorineural deafness. When an electrode is activated, it delivers a series of biphasic pulses, normally with phase durations of 25 µs and an 8-µs inter-phase gap.

The strategy embedded in the sound processor determines which combinations of electrodes to stimulate according to the acoustic signal received by the microphone. Figure 1 shows a simplified block diagram of the cochlear implant. The most commonly used strategy divides the incoming sound signal into as many frequency bands as there are electrodes (22 in the Cochlear Ltd Nucleus devices), selects a small number of the bands with the highest amplitude (typically the 8 highest of the total 22 available), and then stimulates those electrodes at a current level related to the smoothed amplitude in each band.

If a high-frequency pure tone is played, about 1400Hz for example, electrode #13 is stimulated (see Table 1). The audiologist may change the frequency allocations of each electrode individually. However, in a typical clinical session, the allocation will only be changed in case of dysfunctional electrodes.

SPEECH SIGNALS

Speech signals convey semantic meaning though a rapid succession of vowel and consonant sounds. Vowel sounds (such as the /ɪ/ in 'heed', /ɛ/ in 'head', /æ/ in 'had', /ʌ/ in 'hud', /ɒ/ in 'hod', /ɔ/ in 'horde', /ʊ/ in 'hood', /u/ in 'who'd') are produced without significant constrictions in the vocal tract, and are generally "voiced" – that is the vocal folds vibrate and produce a harmonic sound. Vowels in English and other non-tonal languages generally have a fairly consistent voicing frequency (F0), with a unique pattern of harmonics called formants, labelled F1, F2, F3 etc. Depending on the vowel sound produced, the first formant in English varies between 300-770 Hz, and the second between 900-2300 [8, 9]. Most vowel sounds can be distinguished by the first and second formants alone [10].

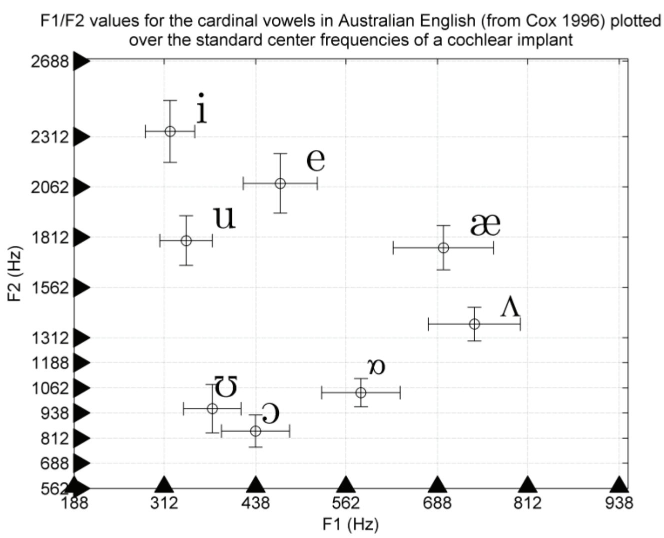

Figure 2. Eight vowels plotted according to the frequency of their first and second formants (F1, F2). The grid overlayed corresponds with the edge of the electrode frequency bands specified in a default CI map (see Table 1).

In a CI, the steady first formant activates one or more of the lowest electrodes, and a number of the higher electrodes are activated by the higher formants – with a different pattern activated for each vowel sound. Thus the CI user receives a fairly unique pattern of electrode activation for each vowel sound. Figure 2 shows the F1 and F2 frequencies for some of the cardinal vowel sounds in Australian English with a male speaker [11], overlaid on a grid representing the edges of the default CI frequency bands. For example the vowel sound "i" has a first formant around 320 Hz and a second around 2320 Hz. It will therefore activate mostly electrodes #22, #21, #10 and #9. Crucially, speech can be understood with a relatively small number of vowel sounds, so that despite the problems of overlapping filter bands and current spread, there is enough frequency resolution using 22 electrodes for many CI users to successfully distinguish between many of the vowels [12, 13]. It is therefore important to allocate frequency bands to electrodes in order to maximise the difference between vowels. As shown in Table 1, a typical frequency allocation table serves that purpose. The low-frequency bands (#22 to #14) increase linearly with a fixed width of 125 Hz. The bandwidth of the higher frequency bands increases logarithmically and reaches 1 kHz for the highest band (#1). The lowest 5 electrodes (inserted deeply) are associated with the first formants of most vowels, the middle 8 electrode to the second formants, and the highest to the third formants and other high pitched sounds.

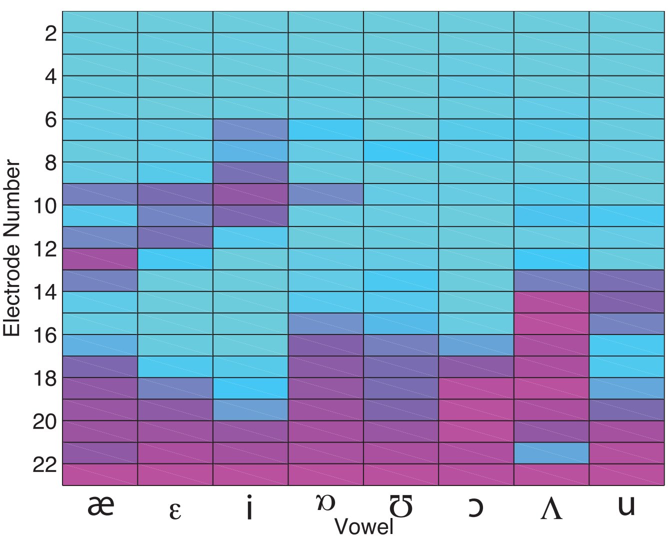

Figure 3. shows the activation pattern of 8 Australian vowels averaged across time. The ordinate represents the average activation of each of the 22 electrodes. The colour codes the amplitude of the activation.

Figure 3 shows that each vowel produces a specific pattern of activation across electrodes that can be learned by the CI recipient. The vowel /ɔ/ will mostly activate the lower frequency electrodes (22 to 17). The vowel /ɪ/ clearly shows the activation of two zones corresponding to the two formants. Compared to vowel sounds, which are distinguishable mostly on the basis of spectral information, consonants are mostly distinguishable on the basis of how the overall amplitude varies in time [14]. The rate of stimulation pulses in CIs can vary from around 200 Hz up to 1200 Hz. At these rates, gross temporal cues can be transmitted fairly well. Thus, despite the complex acoustic nature of consonant sounds, many of the time-based cues used to distinguish between consonant sounds are successfully transmitted to the listener [15].

MUSICAL PITCH

Most musical sounds share these same basic features; spectral parameters encode pitch, melody, and tonal aspects of timbre while time-varying parameters encode rhythm and impulsiveness aspects of timbre. However, musical signals are acoustically more complex than speech. Unfortunately, the signal processing employed in most standard CIs destroys many of the acoustic parameters in the signal, only passing the smoothed amplitude envelopes of a series of band pass filters.

The perception of pitch is based on the fundamental frequency (F0) of an acoustic signal. It is not completely clear how pitch is coded in the auditory system, but research so far points to the conjunction of three physiological cues. First, as described in the above section, different auditory nerves are stimulated depending on the frequency of the acoustic signals. Therefore, frequency information can be transmitted to the brain by detecting which auditory nerves have been activated. This cue is called **place coding**. Second, the basilar membrane within the cochlea resonates, and therefore triggers the auditory nerves at a rate related to the input frequency (at least up to about 1-4 kHz). This temporal pattern of neural firing can also convey pitch information. This is called **temporal coding**. Third, as the high frequencies excite a portion of the membrane located at the entrance of the membrane, and the low frequencies a portion at the end of the membrane, the delay of excitation will be different according to the frequency – the low frequencies will be delayed by the time needed to travel along the cochlea. Therefore, pitch information can also be conveyed through the timing of activation of the nerves (the high frequencies will arrive first). For example, the delay of a 200 Hz tone is about 6-ms longer than the delay of an 8 kHz tone [16]. This is called **phase coding**.

In current sound processing strategies, pitch information is, for the most part, conveyed by temporal cues (amplitude modulation) and place coding, as different electrodes are activated according to the frequency. It might be possible in the future to introduce more electrodes; however, due to the spread of current, it is unclear if this will improve the frequency resolution [17]. Furthermore, as the frequency allocations are not designed to convey musical pitch, the electrical output of the sound processor cannot accurately reproduce a musical scale. Figure 4 shows the activation pattern of all 22 electrodes for a piano note as a function of its fundamental frequency. As the lowest frequency band (electrode 22) ranges from between 188 and 313 Hz, the fundamental frequency of any note one octave below Middle C will not be transmitted. The activation will only be caused by higher harmonics. When the fundamental frequency reaches this octave, the fundamental frequency will start to activate the lowest electrode. Therefore, an increase in the lower musical note will not always be associated with an increase in position of the electrodes activated. This will negatively impact the perception of lower pitch notes.

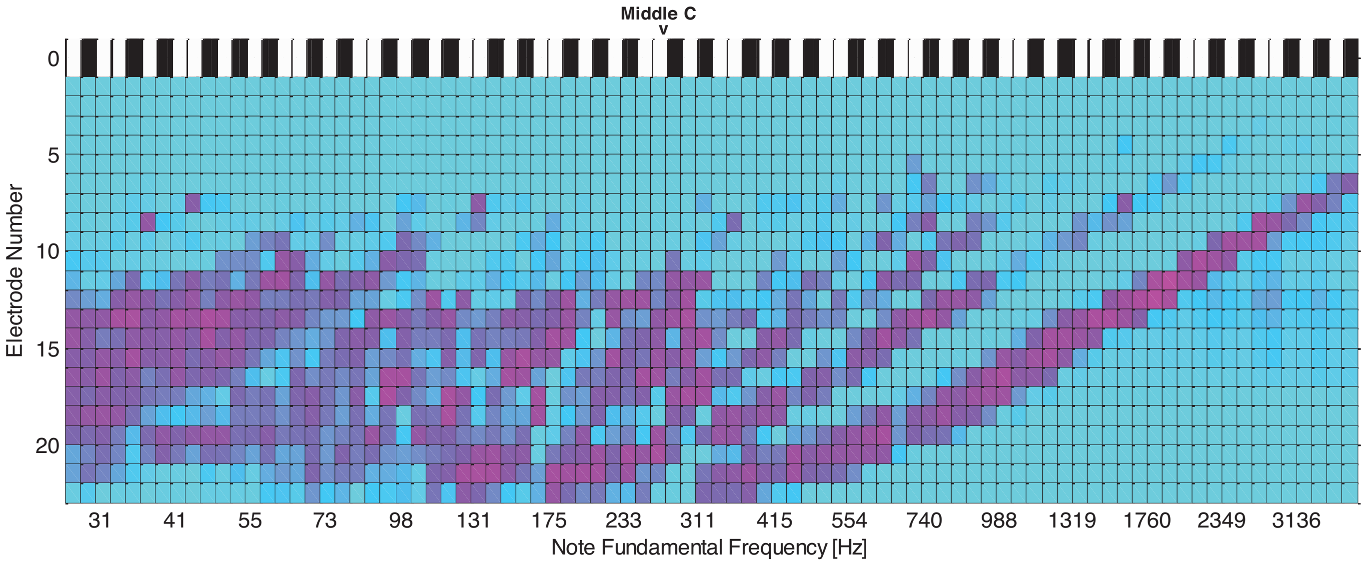

Figure 4. Average electrical excitation diagram of a piano note as function of the fundamental frequency (in Hz on bottom axis, or as piano key top axis). Red, hot colour represents a strong activation, and blue, cold colour, no activation.

In most current CI recipients, the pulse rate is fixed at 900 Hz and therefore amplitude modulation for frequencies lower than half the pulse rate can be transmitted. Some CI recipients use this temporal cue to perceive the pitch of the lower octaves of the piano, where no place cues are available (as showed in Figure 4) [3]. It is possible to increase the pulse rate, however this does not appear to improve pitch perception [18], but does decrease battery life.

Finally, in most current strategies the phase delay is not implemented, so recipients cannot benefit from this cue. Experimental strategies have been tested to determine whether the addition of a phase delay will improve speech perception. Results have found small but significant improvement for speech perception in noise [19].

In summary, CIs only partially convey two out of the three main pitch cues. This explains their poor results in pitch discrimination tasks. A study has shown that most CI recipients could not reliably identify a pitch direction change below three semi-tones or that only 20% could identify a well known melody without rhythm cues [20].

When comparing Figure 3 and Figure 4, it is clear that the frequency allocations of the bands are well suited for speech: every vowel produces a specific pattern of the activation. On the other hand there is no clear increase in activation pattern with successive musical notes below Middle C.

THE INTERACTION BETWEEN SPEECH AND MUSIC IN COCHLEAR IMPLANT RECIPIENTS

We have argued that the strong focus on speech in the development of the cochlear implant was detrimental to the perception of music. However, in some cases speech signals have been shown to enhance music perception.

Experiments using familiar musical items revealed that verbal cues increase CI users' ability to positively identify the musical material [21-24]. Indeed, it was reported that some CI listeners were able to extract linguistic information from sung lyrics [25] and correlations were found between melody recognition with lyrics and speech perception [21].

On the other hand, vowel identity and pitch can be conveyed by the same cues (electrode position), therefore, speech signals could have a detrimental effect on music. Vowels sung to the same pitch should be distinguishable through their differences in spectral shape. As the formants of each vowel will activate different electrodes, this might be perceived as a difference in timbre. For example, a transition from the word "head" to "hod" might be confused with an increase of pitch instead of a change of vowel.

CIs and hearing-aids were developed to assist their users with speech perception and understanding. Advanced sound processing algorithms now include a sound classification system for the automatic recognition of the acoustic environment [26]. Hearing devices are now programmed to automatically distinguish between sound classes such as 'clean speech', 'speech in noise', 'noise', and 'music'. When a speech in noise situation is detected, the algorithms supress the noise and amplify speech signals. Unfortunately, these automatic algorithms can often mis-classify music as speech in noise. This can lead to an inappropriate processing of music.

STRATEGIES DEDICATED TO IMPROVE MUSIC PERCEPTION IN COCHLEAR IMPLANT

As a cochlear implant that can propose reliable speech and music will be an important commercial success, many new sound processors strategies have been tested [6]. Some strategies tried to improve the amount of temporal information conveyed by enhancing the amplitude modulation of each electrode [for example 27, 28], some others by adapting the rate of stimulation to the F0 of the signal on each electrode [for example 29, 30] or by increasing the overall rate of stimulation (such as the HiRes strategy of Advanced Bionics or FSP strategy of Medel devices). Unfortunately none of those strategies succeed in bringing music perception to a satisfactory level [6].

CONCLUSIONS

Although the cochlear implant has restored the communication abilities for hundreds of thousands of people around the world, it has short-comings in music perception. Given the complexity and very restricted size of the cochlea, no hearing device will be able to restore hearing perfectly, at least not with the current technology. It is natural that speech will be the primary focus while developing the cochlear implant. However, a cochlear implant that will be designed specifically for music perception might be possible, but can be unfavourable to speech perception.

ACKNOWLEDGEMENTS

The Bionics Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program. Ninia Simon was financially supported by a scholarship from the German Academic Exchange Service. Portions of this material were published in the Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM); Innes-Brown H, Au A, Stevens C, Schubert E & Marozeau J (2012). New music for the Bionic Ear: An assessment of the enjoyment of six new works composed for cochlear implant recipients, where HIB's attendance was funded by the Harold Mitchell Travelling Fellowship 2012.

REFERENCES

[1] Ramsden, R.T., History of cochlear implantation. *Cochlear Implants International*, 2013. **14**(S4): p. 3-5.

[2] Galvin, J.J., Q.J. Fu, and R.V. Shannon, Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, 2009. **1169**(1): p. 518-533.

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

[4] Looi, V. and J. She, Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *International journal of audiology*, 2010. **49**(2): p. 116-128.

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

[6] McDermott, H., Music Perception, in Auditory Prostheses: New Horizons, A.N.P. Fan-Gang Zeng, Richard R. Fay, Editor 2011, Springer: New York. p. 305-339.

[7] Greenwood, D.D., A cochlear frequency□position function for several species—29 years later. *The Journal of the Acoustical Society of America*, 1990. **87**(6): p. 2592-2605.

[8] Hillenbrand, J., et al., Acoustic characteristics of American English vowels. *J Acoust Soc Am*, 1995. **97**(5 Pt 1): p. 3099-111.

[9] Cox, F., The acoustic characteristics of/hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, 2006. **26**(2): p. 147-179.

[10] Pols, L.C., L.J. van der Kamp, and R. Plomp, Perceptual and physical space of vowel sounds. *J Acoust Soc Am*, 1969. **46**(2): p. 458-67.

[11] Cox, F., Vowel change in Australian English. *Phonetica*, 1996. **56**(1-2): p. 1-27.

[12] Eddington, D.K., Speech discrimination in deaf subjects with cochlear implants. *Journal of the Acoustical Society of America*, 1980. **68**(3): p. 885-91.

[13] Blamey, P.J., et al., Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors. *Journal of the Acoustical Society of America*, 1987. 82(1): p. 48-57.

[14] Diehl, R.L., A.J. Lotto, and L.L. Holt, Speech perception. *Annual review of psychology*, 2004. **55**: p. 149-79.

[15] Shannon, R.V., et al., Speech recognition with primarily temporal cues. *Science*, 1995. **270**(5234): p. 303-4.

[16] Donaldson, G.S. and R.A. Ruth, Derived band auditory brain□ stem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects. *The Journal of the Acoustical Society of America*, 1993. **93**(2): p. 940-951.

[17] Shannon, R.V., Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hearing Research*, 1983. **12**(1): p. 1-16.

[18] Vandali, A.E., et al., Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and hearing*, 2000. **21**(6): p. 608-24.

[19] Taft, D.A., D.B. Grayden, and A.N. Burkitt, Across-frequency delays based on the cochlear traveling wave: enhanced speech presentation for cochlear implants. *IEEE transactions on bio-medical engineering*, 2010. **57**(3): p. 596-606.

[20] Gfeller, K., et al., Accuracy of Cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing*, 2007. **28**(3): p. 412-423.

[21] Leal, M.C., et al., Music perception in adult cochlear implant recipients. *Acta Otolaryngol*, 2003. **123**(7): p. 826-35.

[22] Fujita, S. and J. Ito, Ability of nucleus cochlear implantees to recognize music. *Ann Otol Rhinol Laryngol*, 1999. **108**(7 Pt 1): p. 634-40.

[23] Looi, V. and J. She, Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *Int J Audiol*, 2010. **49**(2): p. 116-28.

[24] Gfeller, K., et al., The effects of musical and linguistic components in recognition of real-world musical excerpts by cochlear implant recipients and normal-hearing adults. *J Music Ther*, 2012. **49**(1): p. 68-101.

[25] Gfeller, K., et al., The impact of voice range and instrumental background accompaniment on recognition of song lyrics. *7th Asia Pacific Symposium on Cochlear Implants and Related Sciences*, 2009: p. 42-46.

[26] Allegro, S., M. Büchler, and S. Launer. Automatic sound classification inspired by auditory scene analysis. in *Proc. of EUROSPEECH 2001, 7th European Conference on Speech Communication and Technology*, 2001, Aalborg, Denmark

[27] Laneau, J., J. Wouters, and M. Moonen, Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neurotology*, 2006. **11**(1): p. 38-52.

[28] Vandali, A.E., et al., Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies. *Journal of the Acoustical Society of America*, 2005. **117**(5): p. 3126-38.

[29] Grayden, D.B., et al. A cochlear implant speech processing strategy based on an auditory model. *Proc. of Intelligent Sensors, Sensor Networks and Information Processing Conference* 2004. Melbourne.

[30] Li, X., et al., Improved perception of music with a harmonic based algorithm for cochlear implants. Neural Systems and Rehabilitation Engineering, *IEEE Transactions* on, 2013. **21**(4): p. 684-694.