Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Ability of cochlear implant users to segregate a melodic line

Jeremy Marozeau (1), Hamish Innes-Brown (1), David B. Grayden (2,1), Anthony N. Burkitt (2,1) and Peter Blamey (1,2)

(1) The Bionic Ear Institute, East Melbourne, Australia
(2) The University of Melbourne, Australia

PACS: 43.64.Me, 43.66.Mk, 43.25.Nm, 43.75.Cd, 43.66.Ts

ABSTRACT

The aim of this study was to examine the effects of four acoustic parameters on the difficulty of extracting a simple 4-note melody from a background of distracter notes. Melody extraction difficulty ratings were recorded while four acoustic parameters of the distracter notes were varied separately: fundamental frequency (F0), intensity, temporal envelope and spectral envelope. The average difficulty ratings for listeners with cochlear implant (CI) users (N=12) were compared with two other groups with normal hearing: with musical training (N=18) and without musical training (N=19). Results show significantly lower results ($p<0.05$) for the CI group compared to the musician group for the F0 and spectral conditions. These results likely reflect the operation of the CI sound processor, which presents gross spectral and temporal envelope cues well, but does not resolve individual harmonics of the fundamental frequency (F0) or fine timing cues.

INTRODUCTION

Music is often composed of different melodic lines that are played together, either on the same or different instruments. These melodic lines, or “streams”, are often defined or separated by a number of perceptual parameters, such as pitch, timbre or loudness (reviewed by Bregman, 1994). One important aspect of listening to music is to be able to hear these melodic lines separately and in comparison with each other. A similar situation arises in social gatherings, where it may be difficult to understand the speech of a specific speaker because one has to segregate their voice from other voices with similar perceptual qualities.

The greater difficulty of listeners with hearing impairment compared to normally hearing listeners when perceiving speech in background noise may be caused by a degraded ability to segregate auditory streams. For hearing-impaired listeners, a larger acoustic difference between sources is required before those sources can be segregated. The increase required for each type of perceptual cue (pitch, timbre, loudness, etc), and how this differs with each hearing device (cochlear implant, hearing aid) is currently unknown.

Very few papers have reported studies on auditory streaming in cochlear implant (CI) users. The ability of CI users to fuse two stream together was measured by Hong & Turner (2006). Half of the CI users performed within the range of the NHL group, and the other half showed less obligatory streaming. Chatterjee *et al.* (2006) asked 7 CI users to indicate whether stimulation on two alternating electrodes (A and B) was perceived as two separate streams. The electrode position of A was fixed and B was varied within the experiment. Only one CI user out of seven showed behaviour consistent with auditory streaming in normal-hearing listeners. It was not clear whether the other six CI users were able to experience the two stimuli as perceptually different (above the just-noticeable difference) and form a single stream. In order to investigate this hypothesis, Cooper & Roberts (2007) asked eight listeners to indicate whether alternating stimulation on two electrodes (ABA_ABA) was perceived as fused or segregated. They found: no bistable perception, as would be expected with normal-hearing listeners, and stream segregation results for most individuals were correlated with their result from a pitch discrimination control task. The authors concluded that CI users showed little or no evidence of automatic stream segregation. They confirmed this conclusion in two follow-up studies (Cooper and Roberts, 2009a, b).

In summary, hearing impairment, particularly using a cochlear implant, reduces the perceptual differences between auditory sources, thereby reducing auditory stream segregation and affecting the ability to enjoy music. Cochlear implant users are known to have poor perception of pitch and timbre but relatively good perception of coarse temporal sound features, such as rhythm.

This study evaluated the ability of participants to separate a melody from a background distracter melody based on four different acoustical parameters and tested the effect of musical training and hearing impairments. Unlike the previous studies that used electrical pulsatile stimuli presented directly to the CI, this study used acoustic stimuli presented from a loudspeaker through the sound processor to maximise the ecological validity of the data. Furthermore, a variety of streaming cues were tested with the same participants.

METHOD

Listeners:

Fourty-nine adults participated in this study. On the same day, each listener participated in a complementary study on the effect of vision on auditory streaming ability (see companion paper Innes-Brown et al., 2010). Listeners were divided into three groups according to their hearing impairment and musical training.

1] Group MUS: The first group was composed of 18 adults (8 male, 10 female) with normal hearing and musical training. Normal hearing was defined as audiometric thresholds less than 20 dB HL (ANSI, 1996) at octave frequencies from 250 to 8000 Hz. A listener was categorised as musically trained based on a hierarchical clustering analysis designed to maximise the group differences on four normalised musical activity variables: 1] sight-reading ability self-ratings, 2] general musical aptitude self-ratings, 3] the number of hours of musical practice per week, and 4] years of musical training (for more information see Innes-Brown et al., 2010). Average age was 31 years old with a standard deviation of 7.2 years.

2] Group NonMus: The second group was composed of 19 adults (9 male, 10 female) with normal hearing and minimum musical training. Average age was 32.2 years old with a standard deviation of 7.9 years.

3] Group CI: The third group was composed of 12 adults using a cochlear implant (6 male, 6 female). They all used a Nucleus® cochlear implant, and a fixed rate sound processor strategy (ACE or SPEAK). Those two strategies have shown remarkable performance for speech recognition in quiet, but have limitations in noisy environments.

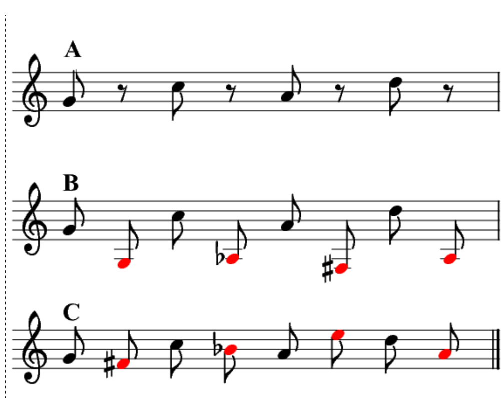

Fig 1: *The 4-note target melody (A), the target, in black, with non-overlapping distracter notes, in red (B), and the target with overlapping distracter notes (C).*

Stimuli:

Two types of sequences were used: the *target was a repeating melody* and the *distracter was pseudo-random notes*. The two sequences will be termed the *target* and the *distracter* in the rest of this document. The two sequences were presented interleaved to the participants (see Fig. 1).

The *target* was a 4-note repeating melody with the following physical parameters: **1] The F0** of the four notes were: 392, 523, 440 and 587 Hz. These frequencies correspond to G, C, A, and D in musical notation (see Fig. 1a). This melody is composed of intervals large enough to be perceived by many people with poor pitch discrimination while being small enough for the notes to be grouped into a single melody (instead of 2 interleaved melodies composed of the 2 low notes and the 2 high notes). **2] The amplitude** of each note of the *target* was adjusted in order to reach the loudness level of 65 Phons (i.e., as loud as a 1-kHz tone at 65 dB SPL) according to loudness models (Glasberg and Moore, 2002, ANSI, 2007). **3] The temporal envelope** of each note was composed of a 30 ms raised-cosine onset, 140ms sustain and 10 ms offset, for a total duration of 180 ms or an impulsiveness of 160 ms, defined as the full duration of the sound at half of the maximum amplitude, FDHM (see Fig. 2). **4] The spectral envelope** of each note consisted of 10 harmonics, successively attenuated by 3 dB.

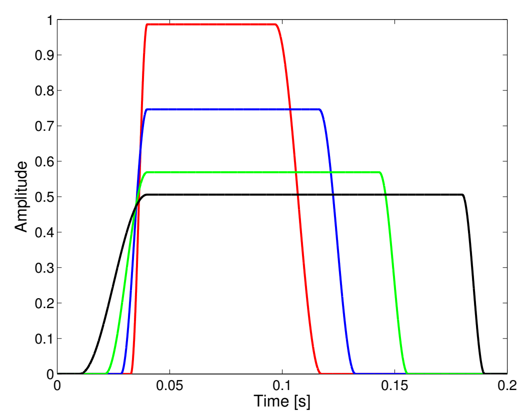

Fig 2: *Different temporal envelopes that vary in impulsiveness defined as the as the full duration of the sound at half of the maximum amplitude, FDHM : black 160 ms, green 119 ms, blue 90 ms and red 70 ms.*

The *distracter* notes consisted of uniformly randomized notes across a range of an octave (absolute range depended on conditions). The distracter notes were varied on one parameter in four different conditions. The parameters were varied gradually within each condition from a level that was designed to induce segregation (see Fig 1B for an example based on the F0 parameter) to the same level as the *target* (Fig 1C). In the latter case, the target and the distracter shared the same physical characteristic, and should be perceived as one fused sequence. The four conditions were: **1] Fundamental frequency.** The 12-semitone F0 range of the possible distracter notes was gradually varied in twenty 1-semitone steps from at least one octave below the target (110 to 208 Hz, i.e. 45 to 56 midinote) to a range that totally overlapped the target (330 to 622 Hz, 64 to 75 midinote). **2] Amplitude.** The amplitude of each distracter note was varied in twenty 2-Phon steps in order to set the loudness from 27 to 65 Phons. **3] Temporal envelope.** Twenty levels of distracter impulsiveness were tested, logarithmically spaced between 60 and 160 ms (see Fig 2). **4] Spectral envelope.** The amplitude of each successive harmonic of the distracter was decreased by the same amount in dB ranging from 25 dB attenuation per harmonic to 3 dB.

The stimuli were constructed using Matlab 7.5 and generated using MAX/MSP 5 through an M-AUDIO Firewire 48-kHz 24-bit sound card. The stimuli were played from a loudspeaker (Genelec 8020APM, selected for its flat frequency response) positioned on a stand at the listener's ear height, 1 m from the listener's head. The experimental protocol was approved by the Human Research Ethics Committee of the Royal Victorian Eye & Ear Hospital.

Procedures:

The experiment consisted of blocks of trials in which the melody was presented continuously interleaved with distracter notes. The distracter notes started with the parameter at the lowest level, called the INC block (*i.e.* the least similarity between target and distracter), or the highest level, called the DEC block (*i.e.* both melody and distracter share the same physical parameter). In the former case, the melody was easily perceived and, in the latter, both melody and distracter were likely to fuse into a single stream and the melody was no longer perceived. After 10 presentations of the 4-note melody (16 seconds), the parameter level of the distracter was either increased (INC block) or decreased (DEC block) to the next level. The block ended when the parameter level reached either level 19 (INC block) or level 0 (DEC block).

An INC block was always run first as a practice session, with the data from this block discarded. Following the practice, each increasing/decreasing block was run twice, with A-B-B-A/B-A-A-B counterbalancing across participants. In the session, listeners participated in another similar experiment on streaming (see companion paper Innes-Brown et al., 2010).

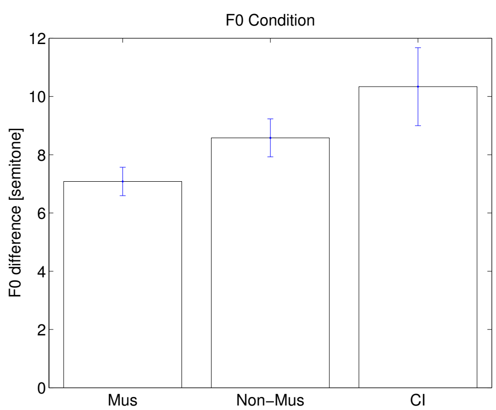

Fig 3: *Average result of the F0 condition for the three groups of listeners. Error bars represent one standard error.*

The listeners continuously rated the difficulty of perceiving the four-note melody using a variable slider on a midi controller (EDIROL U33). The slider was labelled from 0 (no difficulty hearing melody) to 10 (impossible to hear melody). They were instructed to move the slider to the "10" position if the melody was impossible to perceive, and to the "0" position if the melody could be easily perceived. Every time a note of the melody was played, the slider position was encoded in 128 steps on a personal computer using a MAX/MSP 5 environment.

RESULTS

Analysis

For each listener, the four blocks were averaged together and then fitted with a sigmoid function. The knee-point of this function indicated the 50% point of the psychometric function: the level at which the listener rated the difficulty to perceive the melody at mid-point between "*impossible to hear the melody*" and "*no difficulty to hear the melody*." It was assumed that this point indicated the level of bistability: *i.e.* a distracter presented with a higher level would be perceived as segregated more the 50% of the time. Each condition was analysed separately using a one-way ANOVA with listener groups as independent factor and the average knee-point as dependent factor. Effects were considered as significant for $p<0.05$. When significant, a Tukey HSD post-hoc analysis was performed to assess pairwise differences between groups. P values were adjusted in order to take into account multiple comparisons. Analyses were performed using Matlab 7.8, and R.

Fundamental Frequency Condition

Average results for F0 are shown in Fig 3. The ordinate represents the difference of average fundamental frequency between the melody and the distracter. A difference of zero means that both sequences overlapped completely in terms of note range. After a difference of 10 semitones, no note overlap was present between the two sequences. Results showed that, on average, the musician groups needed 7 semitones difference between the melody and the distracter to reach the psychometric function knee-point, the non-musician group needed 8.5 semitones, and the CI group 10.33 semitones. ANOVA revealed a significant effect of groups ($p=0.027$). A Tukey HSD analysis confirmed that the difference between musicians and CI was significant ($p=0.021$). On the other hand, no difference was found between non-musicians and CI ($p= 0.486$). This lack of significance might be partly due to the small group size and greater variability of the CI group.

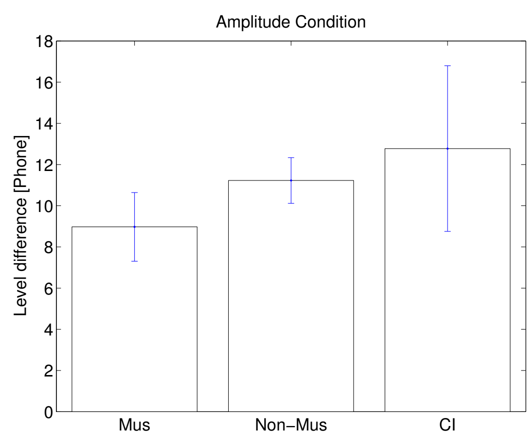

Fig 4: *Average result of the Amplitude condition for the three groups of listeners. Error bars represent one standard error.*

Amplitude Condition

Average results for amplitude are shown in Fig 4. The ordinate represents the difference in Phon level between the melody and the distracter. A difference of zero means that both sequences were presented at the same loudness (65 Phons). A difference of 10 Phons implies that the melody was approximately 2 times louder than the distracter (at 55 Phons). Results show the same pattern as in the F0 condition: the musician group required less level difference (8.97 Phons), followed by the non-musician group (11.22 Phons) and finally the CI group (12.77 Phons). However, because the average differences are small compared to the variance, no significant effect of group was found (p=0.48).

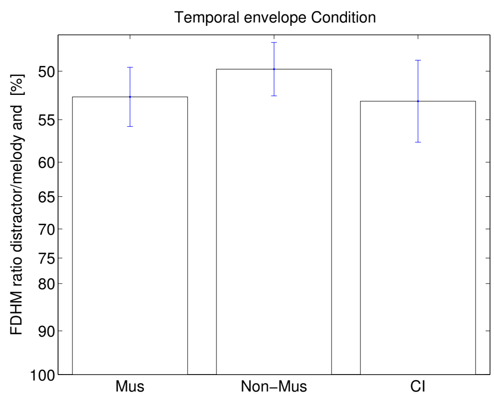

Fig 5: *Average results of the Temporal envelope condition for the three groups of listeners. Error bars represent one standard error.*

Temporal Envelope Condition

Average results for temporal envelope are shown in Fig 5. The ordinate represents the ratio of impulsiveness between the melody and the distracter. A ratio of 100% indicates that both the melody and the distracter share the same temporal envelope. A ratio of 50% indicates that the distracter notes had a temporal envelope half as long as the melody notes (*i.e.* a full duration half maximum amplitude duration of 80 ms). Results show no significant difference between groups (p= 0.72).

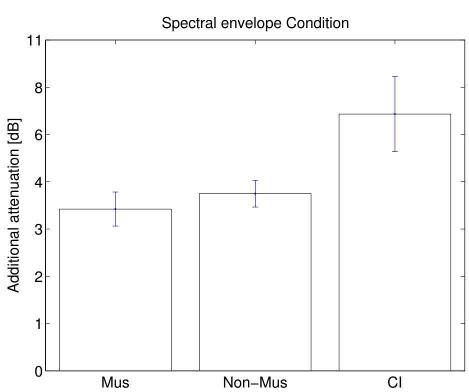

Fig 6: *Average results of the Spectral envelop condition for the three groups of listeners. Error bars represent one standard error.*

Spectral Envelope Condition

Average results for temporal envelope are shown in Fig 6. The ordinate represents the amount of additional attenuation between successive harmonics that were present in the distracter. An additional attenuation of 0 indicates that the distracter was composed of harmonics attenuated by 3 dB (as was the melody). An additional attenuation of 6 dB indicates that the distracter was composed of successive harmonics attenuated by 9 dB. Results show that the musician and non-musician groups needed about the same spectral envelope attenuation to start to segregate the distracter from the melody (about 7 dB of additional attenuation per harmonic). On the other hand, CI users needed a larger difference in the spectral envelope (an additional 11 dB of attenuation per harmonic). ANOVA revealed a highly significant effect of groups (p=0.012). A Tukey HSD analysis confirmed that musicians and nonmusicians showed a lower knee-point than the CI group (p= 0.012, and p=0.039 respectively).

DISCUSSION AND CONCLUSIONS

The average levels required to reach 50% on the psychometric functions were lower for musicians than for non-musicians for all four parameters, reflecting the effect of training on auditory streaming. For CI users, levels were higher when the distracter notes varied in F0 and the spectral envelope. These results reflect the difficulty that CI users have in pitch and timbre discriminations. However, CI users showed levels within the range of normal-hearing listeners when the distracter notes varied in intensity and temporal envelope.

These results likely reflect the operation of the CI sound processor, which presents gross spectral and temporal envelope cues well, but does not resolve individual harmonics of the fundamental frequency (F0) or fine timing cues. The results have implications for the design of new CI sound processors that will enhance music appreciation through the artificial enhancement of specific acoustic cues.

ACKNOWLEDGMENT

The Financial suport were provided by the Jack Brockhoff Foundation; Goldman Sachs JBWere Foundation; Soma Health Pty Ltd; Mr Robert Albert AO RFD RD; Miss Betty Amsden OAM; Bruce Parncutt & Robin Campbell; Winnifred Grassick Memorial Fund. The Bionic Ear Institute acknowledges the support it receives from the Victorian Government through its Operational Infrastructure Support Program.

REFERENCES

ANSI 1996. Specification for audiometers. *In:* AMERICAN NATIONAL STANDARD INSTITUTE (ed.) *ANSI S3.6-1996* New-York: American National Standard.

ANSI 2007. Procedure for the Computation of Loudness of Steady Sounds. *In:* AMERICAN NATIONAL STANDARD INSTITUTE (ed.) *ANSI S3.4-2007.* New-York: American National Standard.

BREGMAN, A. S. 1994. *Auditory Scene Analysis: The Perceptual Organization of Sound,* Cambridge, MA, The MIT Press.

CHATTERJEE, M., SARAMPALIS, A. & OBA, S. I. 2006. Auditory stream segregation with cochlear implants: A preliminary report. *Hear Res,* 222**,** 100-7.

COOPER, H. R. & ROBERTS, B. 2007. Auditory stream segregation of tone sequences in cochlear implant listeners. *Hear Res,* 225**,** 11-24.

COOPER, H. R. & ROBERTS, B. 2009a. Auditory stream segregation in cochlear implant listeners: measures based on temporal discrimination and interleaved melody recognition. *J Acoust Soc Am,* 126**,** 1975-87.

COOPER, H. R. & ROBERTS, B. 2009b. Simultaneous Grouping in Cochlear Implant Listeners: Can Abrupt

Changes in Level Be Used to Segregate Components from a Complex Tone? *J Assoc Res Otolaryngol*.

GLASBERG, B. R. & MOORE, B. C. 2002. A model of loudness applicable to Time-Varying Sounds. *J. Audio Eng. Soc.,* 50**,** 331-342.

HONG, R. S. & TURNER, C. W. 2006. Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *J Acoust Soc Am,* 120**,** 360-74.

INNES-BROWN, H., MAROZEAU, J., GRAYDEN, D. B., BURKITT, A. N. & BLAMEY, P. 2010. Improving musical streaming for cochlear implant users using visual cues. *Proceedings of 20th International Congress on Acoustics*. Sydney, Australia.