

Improving musical streaming for cochlear implant users using visual cues

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

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

PACS: 43.66.MK, 43.66.TS, 43.66.HG

ABSTRACT

The ability to follow separate lines of melody is an important factor in music appreciation. This ability relies on effective auditory streaming, which is much reduced in people with hearing impairment, contributing to their reported difficulties in music appreciation. The aim of this study was to assess whether visual cues could reduce the difficulty of segregating a melody from background notes for 1] people with normal hearing and extensive musical training, 2] people with normal hearing and no musical training, and 3] musically untrained cochlear implant users. Normalhearing musicians (N=18), normal-hearing non-musicians (N=19), and cochlear implant (CI) users (N=12) were asked to rate the difficulty of segregating a four-note repeating melody from random interleaved distracter notes. Visual cues were provided on half the blocks; and difficulty ratings for blocks with and without visual cues were compared between groups. When no visual cues were present, musicians rated the task as less difficult than nonmusicians, with CI users reporting the most difficulty. For normal-hearing listeners, visual cues and musical training both reduced the difficulty of extracting the melody from the distracter notes. However, musical training was not required for the visual cue to be effective, with musically untrained listeners showing the largest reduction in difficulty. CI users also reported significantly reduced difficulty extracting the melody when using the visual cue, reporting similar difficulty ratings to normal-hearing listeners without the aid of the visual cue. These results suggest that visual cues may be an effective means of improving the enjoyment of music for cochlear implant users. Further research is required to optimise the design of the display and to determine the most useful acoustic features for the display to encode.

INTRODUCTION

The appreciation of music is increasingly being recognised as vital to many areas of functioning in society, and has a myriad of beneficial effects on the body and the brain (Gfeller & Knutson, 2003; Mithen, 2009). Music often contains multiple "streams," for instance a melody and a harmony, either played on the same or separate instruments. The ability to separate and group auditory streams is called auditory stream segregation, and this ability is based mainly on acoustic differences (such as pitch and timbre) between the streams. Unfortunately, the sensations of pitch (the "height" of a sound) and timbre (the quality of sound that differentiates instruments) are both degraded by hearing loss, which in turn leads to reduced stream segregation, and reduced appreciation of music. Furthermore, some hearing devices such the cochlear implant (CI) are currently very poor at reproducing music (see Gfeller et al., 2005; and McDermott, 2004 for reviews) and people with hearing impairments may tend to feel excluded in social situations and events where music is present.

Recent work in cognitive neuroscience has found that the sensory modalities are integrated at relatively early stages of processing in the brain (Driver & Noesselt, 2008), and that concurrent stimuli in one sense (vision for instance) can alter

or improve perception in another sense such as audition (Bolognini, Frassinetti, Serino, & Ládavas, 2005; Shams, Kamitani, & Shimojo, 2000). The power of visual cues to improve auditory perception is demonstrated in the case of speech perception in background noise. It has long been known that when a speaker's lip and facial movements are available, an improvement in performance equivalent to increasing the signal-to-noise ratio by up to 15 dB can be observed (Sumby & Pollack, 1954).

In the musical domain there has been little research examining the effect of vision on perception of music, however, concurrent video of musical performances have been shown to affect ratings of tension and phrasing (Vines, Krumhansl,



Figure 1: The simple 4-note melody (G, C, A, D, midinotes 62, 72, 69, 74) depicted on the stave used as the visual display. Each melody note turned red as it played.

Wanderley, & Levitin, 2006), physiological responses to music (Chapados & Levitin, 2008), and the perception of bowing vs. plucking judgements for stringed instruments (Saldaña & Rosenblum, 1993). A concurrent visual cue representing pitch has also been found to improve auditory stream segregation in the context of a classic streaming experiment (Rahne, Böckmann, von Specht, & Sussman, 2007), however it is not known if this improvement can be maintained in a musical task.

People with hearing impairment using cochlear implants have been shown to be better than normally-hearing listeners at integrating visual information with degraded auditory signals (Rouger *et al.*, 2007). If visual information can improve stream segregation in a musical context, people with hearing impairment may also be better able to take advantage of this information. The provision of an appropriate visual cue may thus improve the appreciation of music for users of cochlear implants and hearing aids. Although it may be possible to use such visual information to assist CI users, there is currently very little research on the effect of visual cues on streaming for either normal-hearing listeners or those using cochlear implants, and to our knowledge, none have employed a musically-relevant task.

It is also unknown whether extensive experience or training will be required in order to make use of visual information. In the current experiment, an animated musical stave depicting the melody notes was used as the visual display. In order to test the effect of training, highly experienced musicians, with extensive training associating visual depictions of pitch with their auditory correlates, were also assessed in order to investigate the effect of training.

In this experiment, the effect of visual cues on musical streaming in cochlear implant users, and normal-hearing listeners with and without musical training was examined. A musical streaming paradigm was employed that involved the extraction of a simple melody from a background of distracter notes varying in pitch. The difficulty of extracting the melody was then compared depending on whether or not a concurrent visual cue was present.

METHODS

Ethics statement

The experimental protocol conforms to The Code of Ethics of the World Medical Association (Declaration of Helsinki), and was approved by the Human Research Ethics Committee of the Royal Victorian Eye & Ear Hospital (Project 09-880H). Written informed consent was obtained from all participants involved in the study.

Participants

Forty-nine adults participated, 37 normal-hearing and 12 cochlear implant users. In order to classify the 37 normalhearing participants as musicians or non-musicians objectively, those participants were divided into two groups according to a hierarchical cluster 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. The cluster analysis was constrained to two possible solutions. The group composed of participants with higher scores on the musical evaluation form was designated "Musicians" (MUS: N=18), with the remainder "Non-musicians" (NMUS: N=19). The means and standard deviations of the musical activity variables separated by the results of the cluster analysis are summarised in Table 1, along with ages and gender details for all groups. All normal-hearing participants reported normal hearing and all participants reported normal or correctedto-normal colour vision. Ten of the 12 CI users used the Advanced Combination Encoder (ACE) strategy, and 2 used the spectral-peak (SPEAK) strategy. All used Cochlear Ltd Freedom (N=2) or Nucleus (N=10) cochlear implants and Freedom (N=5), Esprit 3G (N=5) or SPEAR (N=2) sound processors. Travel and lunch expenses were reimbursed AU\$40.



Figure 2: Experimental design. Melody notes (black dots) are repeated continuously. Distracter notes (red dots) are chosen randomly from a pool of possible notes (black square) which slowly increases (in INC blocks) towards the melody notes throughout the block. In DEC blocks, the pattern is reversed.



Figure 3: Average (+/- SEM) difficulty ratings as a function of distracter level for non-musicians (NMUS), musicians (MUS) and co-chlear implants users (CI).

1			
	NMUS	MUS	CI
N(females)	19(10)	18(10)	12(6)
Mean age(SD)	31(7.2)	32.2(7.9)	67.7(9.1)
Sight-reading(SD)	1.6(1.9)	4.4(1.1)	-
Aptitude (SD)	1.0(1.3)	4.3(.8)	-
Hours Prac. (SD)	1.5(3.4)	17.1(10.8)	-
Years Playing (SD)	4.9(5.4)	24.2(6.3)	-

Stimuli

The melody and distracter notes were constructed using Matlab 7.5 and presented using MAX/MSP 5 through an M-AUDIO 48-kHz 24-bit Firewire sound card. Each note consisted of a 180 ms complex tone with 10 harmonics. Each successive harmonic was attenuated by 3 dB, and each note included a 30 ms raised-cosine onset and 10 ms offset. The notes were played from a loudspeaker (Genelec 8020APM) positioned on a stand at the listener's ear height, 1 m from the listener's head. Each note was equalised in loudness to 65 phons according to a loudness model (ANSI, 2007).

The participants were exposed to a series of notes presented every 200 ms. Within this series of notes was a repeated fournote target melody and interleaved distracter notes. The target melody pitches (see Figure 1) were G, C, A, and D above middle C (midinotes 67, 72, 69, and 74 respectively). The melody was composed of intervals large enough to be perceived by people with poor pitch discrimination (as it is often the case in cochlear implant listeners) while being small enough for the sequence to be grouped into a single stream (instead of 2 interleaved streams composed of the 2 low notes and 2 high notes). For convenience, note pitches are referred to throughout using standard midinote values - middle C is designated 'midinote 60', with each integer corresponding to a semitone change in pitch. Each distracter note value was randomly chosen from a pool of 12 consecutive midinotes spanning an octave. Throughout the experiment, the note range of this octave pool was gradually varied providing a range of melody-distracter separation, or overlap levels (see Figure 2 and Procedure section). It is worth noting that as the distracter notes were chosen randomly from every possible midinote within the octave range, so the distracter notes were not necessarily in the same tonality (key) as the melody. However, it has been shown previously (Dowling, 1973), that tonality has little effect on the difficulty of extracting a melody from interleaved background notes.

Procedure

Four counterbalanced sessions were run for each participant – one with the visual cue present (Vision) and one without (Novision). In both Vision and No-Vision sessions, the distracter note range could either slowly increase (INC) or decrease (DEC). An INC block is shown in Figure 2.

In INC blocks, the distracter note range was varied in 20 levels from no overlap (a separation of one octave between the highest distracter note and the lowest melody note) to total overlap. The distracter notes were initially picked from the range of midinotes 45-56. The range of possible distracter notes was then slowly increased until they completely overlapped the melody (midinote range 64 to 75). In each level, the melody was repeated 10 times (lasting 16 seconds). In DEC blocks, the order was reversed.

Before each test session, the melody was presented 20 times without distracter notes; and an INC practice block followed. During testing, each INC/DEC block was repeated twice, with INC-DEC-DEC-INC or DEC-INC-INC-DEC order counterbalanced across participants. The duration of each block was about 5 minutes, and each session lasted about 30 minutes. In order to reduce possible pitch memory effects between Vision and No-vision sessions, a pitch increment, randomly chosen between 0 and 4 semitones, was added to all notes of the same session.

The participants were asked to rate the difficulty of perceiving the four-note melody continuously throughout each block 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). Participants 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.



Figure 4: Difficulty ratings (+/- SEM) for non-musicians (NMUS), musicians (MUS) and cochlear implant users (CI) averaged across all distracter not levels.

RESULTS

The difficulty ratings were averaged across INC and DEC blocks, and across the two repeats of each block. Figure 3 shows the average difficulty ratings as a function of distracter note range level, for Vision and No-Vision blocks in each group. When the distracter note level was low (with an octave separation between the melody and distracter), all participants rated the task as relatively easy. As the distracter note level increased, average difficulty levels increased in a monotonic fashion until the maximum distracter note level, when the distracter notes were completely overlapping the melody notes. At this point, most participants reported the maximum difficulty in extracting the melody notes from the distracters. Figure 4 shows the difficulty ratings averaged across all levels of the distracter. As can be seen in Figure 4, musicians generally rated the task as less difficult that nonmusicians, and cochlear implant users difficulty ratings were generally higher than both the normal-hearing groups, and also showed greater variability.

In order to asses the significance of these effects, the difficulty ratings were entered into a repeated-measures mixed ANOVA with a between-groups factor Group (NMUS, MUS, CI), and within-groups factors for Vision (Vision, No-vision), and distracter Level (20 levels, from complete overlap to one octave separation). Hochberg's GT2 procedure was used to control Type I error rate in post-hoc tests where the group sizes were unequal, and Mauchley's test was used to estimate sphericity. Greenhouse-Geisser corrected p levels and estimates of sphericity (ϵ) are reported if Mauchley's test was violated. There was a significant main effect of Group (F[2,46] = 4.0, p=.02). Post-hoc tests indicated that cochlear implant users reported significantly higher difficulty ratings than the musicians, but not the non-musicians. There were also significant effects of Vision, (F[1,46]=22.6, p<.001) and Level (F[19,874]=488.2, p<.001, $\varepsilon=.14$). There was also trend towards a significant interaction between Vision and Group (F[2,46]=2.6, p=.08). We followed up this borderline significant interaction using pairwise comparisons, and found that while non-musicians showed no significant reduction in difficulty while the visual display was present (p=.3), both non-musicians (p < .001) and cochlear implant users (p = .003) showed highly significant reductions. This can be most clearly seen in Figure 4, where average difficulty ratings across all levels are shown for each group.

CONCLUSIONS

In the present study, it was demonstrated that for both normal-hearing and cochlear implant users, basic visual cues depicting the pitches in a simple melody can reduce the difficulty of extracting the melody from background notes. No special training was required for either normal-hearing or cochlear implant users to make effective use of the visual cues. These results have significant implications for the design of future visual aids that may make music more enjoyable for cochlear implant users.

Pitch. In the current study, difficulty ratings generally increased monotonically as the distracter notes increased in pitch towards the range of the melody notes. This result is in agreement with previous research (Dowling, 1973) examining the ability to segregate melodies from interleaved distracter notes. In Dowling's studies, participants were required to name a familiar melody rather than rate the difficulty of extracting a repeating melody, but the results are similar - when the distracter notes completely overlapped the range of the melody notes, the participants in Dowling's experiment were generally unable to name the familiar melodies. As the interleaved distracter notes decreased in pitch, away from the range of melody notes, participants began to nominate the familiar melodies. A similar pattern was seen in the current study, when participants were unable to segregate the melody while the distracter notes overlapped in pitch.

Vision. Visual information has been previously shown to influence stream segregation (Rahne, et al., 2007). In Rahne et al (2007), the frequency separation and rate of a sequence of high and low tones was chosen so that the perception could either be of one or two streams. A visual stimulus, arranged to complement either the one- or two-stream perception, produced a bias towards the corresponding perception, and influenced mismatch-negativity responses to occasional deviants in the high-low sequence. The effect of visual stimuli on auditory processing has also been described at low levels in the brain. It has been shown that visual cues can improve the encoding of pitch and timbre in the auditory brainstem, particularly in musicians (Musacchia, Sams, Skoe, & Kraus, 2007; Musacchia, Strait, & Kraus, 2008). The improvement in representations of these acoustic features in the brainstem may lead to more salient perceptual differences between sounds, and hence this mechanism could possibly explain the effects of visual stimuli found in Rahne et al (2007) as well as the current experiment. The current results extend these findings to the case of melody segregation, by showing that visual cues can reduce the difficulty of extracting a melody from background notes. Whether the visual effect on streaming is a result of improved encoding of acoustic features in the brainstem, or due to more top-down effects of the visual stimulus, is currently unknown, and a topic for further investigation.

Streaming in CI listeners. Previous research investigating stream segregation using interleaved stimuli in CI users has generally found that streaming is difficult (Chatterjee, Sarampalis, & Oba, 2006; Hong & Turner, 2006), if not impossible (Cooper & Roberts, 2007; Cooper & Roberts, 2009). One of the most intriguing results from the current study was that while CI users did report more difficulty extracting the melody than normal-hearing listeners, their overall performance was better than the previous research would suggest is possible. When the visual cues were present, the grand mean difficulty rating for CI users was not significantly different to normal-hearing listeners without the benefit of the visual cue (Figure 4). Previous research in this area has stressed the methodological importance of limiting the stimuli in the streaming tasks to single electrodes, either via direct stimula-



Figure 5: An electrodogram showing four repetitions of the 4-note melody used in the study. The complex tones used stimulate a specific pattern of electrodes for each note.

tion of single electrodes (Chatterjee, *et al.*, 2006) or by using pure tones with frequencies matched to the centre-frequency of each electrode (Cooper & Roberts, 2009). In the current study however, we were interested in maintaining as much musical validity as possible, and so utilised complex tones, with ten harmonics, presented via loudspeaker in free-field conditions. The pattern across electrodes was thus fairly unique for each note (see Figure 5 for an electrodogram showing melody notes only from a single participant), and might have led to increased perceptual differences between melody and distracter notes. Since the ability to segregate streams is mainly based on perceptual differences between sources, this may have led to an increase in the ability to segregate.

Musicians and training: Musicians undergo an intensive period of training, often lasting a lifetime. This training frequently involves segregating and integrating multiple streams of sound, and for most musicians, involves the repeated association of visual notation with an auditory equivalent. This training has been found to have a variety of effects on behaviour, brain structure and function (Schneider *et al.*, 2002; Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005).

In the current study, the musically-trained participants generally reported less difficulty than untrained participants in extracting the melody from background distracter notes when no visual cues were provided. These results are in agreement with several studies showing improved stream segregation in musicians (Beauvois & Meddis, 1997; Vliegen & Oxenham, 1999; Zendel & Alain, 2009). Previous work has also suggested that musicians use visual information more effectively than non-musicians to represent brainstem-level features of sound (Musacchia, et al., 2007; Musacchia, et al., 2008), and thus it was expected that musicians would gain more from the visual cues in the current experiment. However, musicians reported no less difficulty when visual cues were provided. This finding was unexpected, and cannot be explained by floor effects, as musicians still reported significant difficulty extracting the melody when the melody and distracter overlapped. More research is required to explain this finding. One possibility is that although musicians are very well trained in the auditory aspect of this task, the auditory-visual aspect of this task may have served more as a distraction to what the musicians viewed as purely auditory task.

Conclusion: The current study was undertaken to determine whether the provision of simple visual cues might improve the ability of cochlear implant users to segregate a melody from background notes, and whether training would be required in order to use the cues. It was shown that the provision of these cues could indeed reduce the difficulty of segregating the melody, and cochlear implants users reported no more difficulty in this task than normal hearing participants with no assistance from visual cues. These results suggest that simple visual displays may be useful for the hearingimpaired to improve their enjoyment of music. Further research is required to understand which acoustic cues to encode visually, the specific types of visual cues that are most useful, and whether improvements using these cues will generalise to other listening situations.

ACKNOWLEDGEMENTS

Financial suport was 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. (2007). Procedure for the Computation of Loudness of Steady Sounds: American National Standard.
- Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Percept Psychophys.*
- Bolognini, N., Frassinetti, F., Serino, A., & Ládavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental brain research*, 160(3), 273-282.
- Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. [10.1016/j.cognition.2008.05.008]. Cognition, 108(3), 639-651.
- Chatterjee, M., Sarampalis, A., & Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. [10.1016/j.heares.2006.09.001]. *Hearing Research*, 222(1-2), 100-107.
- Cooper, H. R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. [10.1016/j.heares.2006.11.010]. *Hearing Research*, 225(1-2), 11-24.
- Cooper, H. R., & Roberts, B. (2009). Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition. [10.1121/1.3203210]. The Journal of the Acoustical Society of America, 126(4), 1975-1987.
- Dowling, W. J. (1973). The perception of interleaved melodies. [10.1016/0010-0285(73)90040-6]. Cognitive Psychology, 5(3), 322-337.
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57(1), 11-23. doi: 10.1016/j.neuron.2007.12.013
- Gfeller, K., & Knutson, J. (2003). Music to the impaired or implanted ear. Psychosocial implications for aural rehabilitation. (July 2009). Retrieved from

http://www.asha.org/publications/leader/archives/2003/q2/f030429a.htm

- Gfeller, K., Olszewski, C., Rychener, M., Sena, K., Knutson, J. F., Witt, S., et al. (2005). Recognition of" Real-World" Musical Excerpts by Cochlear Implant Recipients and Normal-Hearing Adults. *Ear and hearing*, 26(3), 237-237.
- Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. [10.1121/1.2204450]. *The Journal of the Acoustical Society of America*, 120(1), 360-374.
- McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. [10.1177/108471380400800203]. *Trends in Amplification*, 8(2), 49-82.
- Mithen, S. (2009). The Music Instinct. [10.1111/j.1749-6632.2009.04590.x]. Annals of the New York Academy of Sciences, 1169(The Neurosciences and Music III Disorders and Plasticity), 3-12.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. [10.1073/pnas.0701498104]. Proceedings of the National Academy of Sciences, 104(40), 15894-15898.
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. [10.1016/j.heares.2008.04.013]. *Hearing Research*, 241(1-2), 34-42.
- Rahne, T., Böckmann, M., von Specht, H., & Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. [10.1016/j.brainres.2007.01.074]. *Brain Research*, 1144, 127-135.
- Rouger, J., Lagleyre, S., Fraysse, B., DeNeve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlearimplanted deaf patients are better multisensory integrators. [10.1073/pnas.0609419104]. Proceedings of the National Academy of Sciences of the United States of America, 104(17), 7295-7300.
- Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, 54(3), 406-416.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. [10.1038/nn871]. Nat Neurosci, 5(7), 688-694.
- Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005). Structural, Functional, and Perceptual Differences in Heschl's Gyrus and Musical Instrument Preference. [10.1196/annals.1360.033]. Annals of the New York Academy of Sciences, 1060(The Neurosciences and Music II: From Perception to Performance), 387-394.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, 408(6814), 788.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustic Society of America*, 26, 212-215.

Proceedings of 20th International Congress on Acoustics, ICA 2010

- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. [10.1016/j.cognition.2005.09.003]. Cognition, 101(1), 80-113.
- Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*, *105*, 339-339.
- Zendel, B. R., & Alain, C. (2009). Concurrent Sound Segregation Is Enhanced in Musicians. [10.1162/jocn.2009.21140]. J Cogn Neurosci, 21(8), 1488-1498.