The Second International Conference on Music Communication Science, 3-4 December 2009, Sydney, Australia
http://marcs.uws.edu.au/links/ICoMusic09/index.html

VISUAL INFORMATION CAN IMPROVE MUSICAL STREAM SEGREGATION.

Hamish Innes-Brown[1], Jeremy Marozeau[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

ABSTRACT

Hearing impairment often leads to reduced auditory stream segregation, and hence reduced appreciation of music. An important factor in the appreciation of music is the ability to separate melody from harmony, or melodies played by different instruments. Enjoyment of music is an important part of life that may be difficult for people with hearing impairments, especially those using cochlear implants or hearing aids. Recent work in cognitive neuroscience has suggested that the integration of auditory and visual information in the brain occurs at earlier processing stages than once thought, and that sensory ambiguities can be resolved through this process. The aim of this study was to assess whether a visual cue could improve the segregation of a simple melody from a background of distracter notes. The distracter notes varied in pitch throughout the experiment providing a range of difficulty. The results showed that the visual cue reduced the difficulty of extracting the melody across a wide difficulty range. These results suggest that it may be possible to use a visual display to improve music appreciation for people with impaired hearing.

1. 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 [1]. Music often contains multiple “streams”, for instance a melody and a harmony, either played on the same or separate instruments. The ability to group and separate auditory streams is called auditory stream segregation and is based mainly on acoustic differences (such as pitch and timbre) between the streams. Unfortunately, the sensations of pitch (the “height” of a sound on a scale) 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 [2, 3] 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 [4], and that concurrent stimuli in one sense (vision for instance) can alter or improve perception in another sense (audition) [5, 6]. 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 [7].

In the musical domain there has been little research examining the effect of vision on perception of music; however, concurrent videos of musical performances have been shown to affect ratings of tension and phrasing [8], physiological responses to music [9] and the perception of bowing vs. plucking judgements for stringed instruments [10]. A concurrent visual cue representing pitch has also been found to improve auditory stream segregation in the context of a classic streaming experiment [11]; 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 [12]. 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, and the provision of an appropriate visual cue may 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 normal-hearing listeners, and to our knowledge, none have employed a musically-relevant task.

In this experiment, the effect of visual cues on musical streaming in normal-hearing listeners 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.

2. METHODS

Participants

Twelve participants (7 females and 5 males) were recruited from the community using social networks. The mean age was 35 years (SD 4.3, range 30-45). All participants reported normal hearing and normal or corrected-to-normal colour vision. All participants were administered a short musical experience questionnaire. They rated themselves on a 0-5 scale for sightreading ability and general musical aptitude, and we recorded the number of hours practice per week. All participants rated 3 or less on the sight-reading scale and were not currently practicing or performing any musical instrument. No one was paid for their participation but travel expenses were reimbursed. The experimental protocol was approved by the Human Research Ethics Committee of the Royal Victorian Eye & Ear Hospital.

Stimuli

All notes consisted of 180 ms complex tones with 30 ms rise/fall times. The sounds had 10 harmonics, with each successive harmonic attenuated by 3dB. The sounds were constructed using Matlab 7.5. A repeated four-note target melody was presented with interleaved distracter notes. The target melody (see Fig 1) was made up of midinotes 67, 72, 69 and 74 (with F0 of 392, 523, 440, 587 Hz respectively). For each presentation of a distracter note, the midinote value was randomly chosen from a pool of 12 midinotes spanning an octave. Throughout the experiment, the note range of this octave pool was gradually varied, thus changing the pitch of the distracter notes (as described in procedure). The notes were played from a loudspeaker (Genelec 8020APM) positioned on a stand at the listener's ear height, 1m from the listener's head.

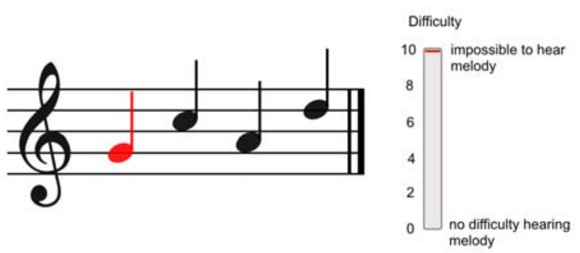

Figure 1: A screenshot showing the visual display (left) and the response indicator (right). Each note of the melody turned red as it played.

The visual display consisted of a musical staff with the 4-note target melody depicted in standard musical notation. Each note in the melody turned red as the melody played. In this way, the visual display depicted the shape of the whole melody, as well as the current note playing. To ensure participants did not have to look down at the slider during the experiment, a visual depiction of the response slider was shown on the screen immediately to the right of the staff. The current position of the slider was updated in real time and shown in red. The visual display is shown in Fig 1. All the sounds were equalised in loudness [13] to 65 phon (equal to a 1 kHz pure tone at 65 dB SPL).

Figure 2: The 4-note melody (A), the melody with non-overlapping distracter notes (B), and the melody with overlapping distracter notes (C).

Procedure

The experiment consisted of blocks in which the melody was presented continuously with the pitch range of the interleaved random distracter notes gradually increasing or decreasing over the five minute length of the block. Blocks were run with the 12-note distracter range either increasing in pitch towards the melody (*INC* blocks) or decreasing away from the melody (*DEC* blocks). In the *DEC* direction, the distracter notes were initially picked from an octave range covering notes from two semitones below the lowest note in the melody to two semitones above the highest note in the melody (completely overlapping the melody). In INC blocks, the distracter notes started with pitches completely separate from the melody, with the highest possible distracter note 11 semitones lower than the lowest melody note. The pitch range of possible distracter notes was then progressively increased in 20 steps of one semitone, with the melody repeated 10 times per level, until the distracter note pitches completely overlapped the melody. Thus, in *INC* blocks, the task of separating the melody started at the easiest level of difficulty and gradually became harder. The melody alone is shown in Fig 2A, the melody with exemplary non-overlapping distracter notes in Fig 2B, and the melody with exemplary overlapping distracter notes in Fig 2C. In *DEC* blocks, the direction was reversed, such that the task started at the hardest level and gradually became easier. 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. The experiment was run twice – once with the visual display (V) and once without (NV). The visual display presentation order was also counterbalanced across participants.

ISBN 978-1-74108-203-6

The listeners 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). The slider position was encoded in 128 steps. Participants were instructed to continuously rate the difficulty of perceiving the target melody on a scale of zero to ten. They were instructed to move the slider to the ten position if the melody was impossible to perceive, and to the zero position if the melody could be easily perceived. When the visual stimulus was present, participants were instructed to concentrate on each note in the visual display as it turned red. In the session, listeners participated in another similar experiment on streaming (see companion paper[14])

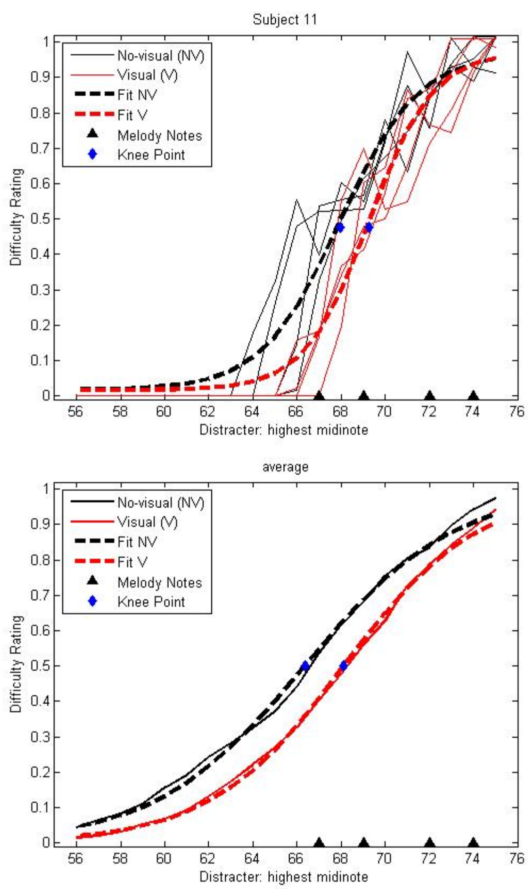

Figure 3: The top panel shows the mean difficulty ratings from a representative individual for each of the 20 distracter pitch levels. Each of the four blocks is plotted separately, with visual blocks in red and non-visual in black. The grand mean difficulty ratings are plotted in the bottom panel. For both plots, sigmoid-fits of the visual and non-visual data are overlaid in dashed lines. The four melody note values are marked with triangles on the x-axes, and the knee points are marked with blue diamonds.

3. RESULTS

The slider position was recorded for each melody note that was played. The data were first normalised to a difficulty scale from zero to one, with a value of zero indicating no difficulty segregating the melody and a value of one indicating the melody was impossible to segregate. An area measure under the data was calculated separately for each block. The mean difficulty ratings for each of the 20 distracter note pitch levels were then calculated separately for each block. A sigmoid function was fitted to the average of these values across the two repeats in each direction (*INC* and *DEC* blocks). The sigmoid function had three degrees of freedom: the amplitude (*a*), the slope (*s*) and the

$$R_{fit} = \frac{a}{1+e^{s(k-L)}},$$

knee point (*k*)

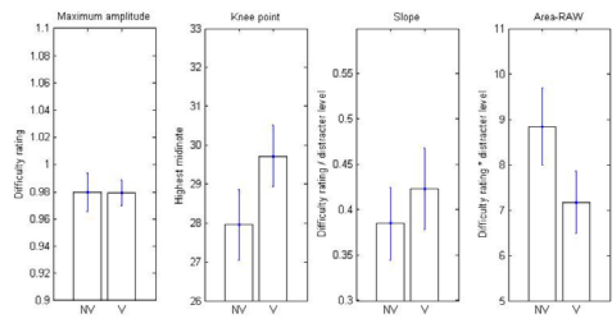

Figure 4: The means and standard errors of the maximum amplitude, slope, knee point and area under the raw data. Values for no-vision (NV) and vision (V) blocks are plotted separately.

where R_{fit} is the predicted response and L is the distracter midinote level (for the highest of the 12-midinote range). The amplitude, *a*, was constrained to values between zero and one.

Figure 3 shows the results from one participant (top panel) and the average across all participants (bottom panel). As can be seen in the average data, participants generally had no difficulty segregating the melody when the distracter notes had midinote values much lower than the melody. As the midinote values increased towards the melody, however, the difficulty ratings began to increase until they approached a maximum value when the distracter values completely overlapped the melody notes. With no visual display, difficulty ratings were about 56% when the distracter notes begin to overlap the melody notes. When the visual display is present, difficulty ratings dropped to about 41%.

ISBN 978-1-74108-203-6

When the visual display was present, the difficulty ratings started to increase at a later point – when the distracter midinote values were higher, or closer to the melody notes. At the 50% difficulty rating point, the distracter notes could overlap the melody notes by an additional two semitones with the same difficulty rating. The difficulty ratings remained lower in visual trials until the highest difficulty levels, where the visual cue appeared to provide less assistance.

In order to test for the effect of the visual display on the difficulty of segregating the melody from the distracter notes, paired t-tests (two-tailed) were performed between visual and non-visual blocks using the three parameters of the sigmoid fit, as well as the mean area under the raw data, for each individual. The means and standard deviations of the four parameters tested are plotted in Figure 4.

Maximum amplitude (*a*): There was no significant difference between the maximum amplitude of the sigmoid fit between vision and no-vision blocks ($t(11) = -0.03$, $p = 0.97$). This indicates that at the point of maximum difficulty, visual cues had no significant effect on the difficulty of extracting the melody.

Slope (*s*): There was no significant difference in the slope between the vision and no-vision blocks ($t(11) = 0.98$, $p = 0.34$). This indicates that the visual cue did not affect the rate of change of difficulty ratings as the distracter notes changed.

Knee point (*k*): There was a significant difference between the knee points of the sigmoid fit between vision and no-vision blocks ($t(11) = 3.4$, $p = 0.005$). This indicates that at the point where difficulty was rated at 50% of the maximum, the distracter notes were significantly closer to the melody notes in the V condition than the NV condition.

Area: There was a significant difference between the mean area under the raw data for vision and no-vision blocks ($t(11) = -4.3$, $p = 0.001$). This indicates that overall the task was easier in the V condition.

4. DISCUSSION

In summary, we have shown that visual cues can be helpful in segregating a simple melody from a background of distracter notes separated by a variable pitch difference. The visual cue decreased melody segregation difficulty ratings by about 15% on average at the point when the distracter notes began to overlap the melody. The visual display had the most benefit throughout the middle range of difficulty, with little benefit when the difficulty was at its highest point. At the highest difficulty levels, however, the distracter notes completely overlapped the melody in pitch, and for most participants the melody was impossible to extract.

The results indicate that an appropriate visual cue may be able to provide enhanced music appreciation for people with impaired hearing by improving their segregation of melodies from other notes. The nature of the visual cue and the training required to interpret it are still unknown, and are the topics of future investigations.

5. REFERENCES

1. Gfeller, K. and J. Knutson (2003) *Music to the impaired or implanted ear. Psychosocial implications for aural rehabilitation.*, http://www.asha.org/publications/leader/archives/2003/q2/f030429a.htm
2. McDermott, H.J., *Music perception with cochlear implants: a review.* Trends Amplif, 2004. **8**(2): p. 49-82.
3. Gfeller, K., et al., *Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults.* Ear and Hearing, 2005. **26**(3): p. 237-50.
4. Driver, J. and T. Noesselt, *Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments.* Neuron, 2008. **57**(1): p. 11-23.
5. Bolognini, N., et al., *"Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs.* Experimental Brain Research, 2005. **160**(3): p. 273-82.
6. Shams, L., Y. Kamitani, and S. Shimojo, *Illusions. What you see is what you hear.* Nature, 2000. **408**(6814): p. 788.
7. Sumby, W.H. and I. Pollack, *Visual contribution to speech intelligibility in noise.* Journal of the Acoustic Society of America, 1954. **26**: p. 212-215.
8. Vines, B.W., et al., *Cross-modal interactions in the perception of musical performance.* Cognition, 2006. **101**(1): p. 80-113.
9. Chapados, C. and D.J. Levitin, *Cross-modal interactions in the experience of musical performances: physiological correlates.* Cognition, 2008. **108**(3): p. 639-51.
10. Saldaña, H.M. and L.D. Rosenblum, *Visual influences on auditory pluck and bow judgments.* Perception & psychophysics, 1993. **54**(3): p. 406-16.
11. Rahne, T., et al., *Visual cues can modulate integration and segregation of objects in auditory scene analysis.* Brain Research, 2007. **1144**: p. 127-35.
12. Rouger, J., et al., *Evidence that cochlear-implanted deaf patients are better multisensory integrators.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(17): p. 7295-300.
13. ANSI, *Procedure for the Computation of Loudness of Steady Sounds*. 2007, American National Standard.
14. Marozeau, J., et al., *The effect of temporal envelope on melody segregation*, in *The Second International Conference on Music Communication Science*. 2009: Sydney.