

Preliminary Results on Spectral Shape Perception and Discrimination of Musical Sounds by Normal Hearing Subjects and Cochlear Implantees

Thomas H. Stainsby (1)
 Hugh J. McDermott (2)
 Colette M. McKay (3)
 Graeme M. Clark (4)

Department of Otolaryngology, University of Melbourne, Australia

(1) stainsbyt@mail.medoto.unimelb.edu.au

(2) hugh@mail.medoto.unimelb.edu.au

(3) mckayc@mail.medoto.unimelb.edu.au

(4) clarkg@mail.medoto.unimelb.edu.au

Abstract

This paper presents an overview of an ongoing research project investigating the perception of musical timbre by people with normal hearing, impaired hearing, and cochlear implants. The investigation of musical timbre has been limited to the perception of steady-state frequency spectra from 10 different sources, including sampled acoustic instruments, sung vowels, and synthetic waveforms. Subjects were tested in three different tasks: 1) the discrimination of spectra when presented in all possible pairs; 2) the measurement of the internally-perceived frequency spectra using a forward-masking paradigm; and 3) the identification of the spectra by name with the restricted set of sound sources from which they were sampled. Preliminary results from the normally-hearing subjects show the spectra to be 99.8% distinguishable, and that significant detail is evident in the internal spectral envelopes from different sounds. There was around 50%-correct identification of stimuli by name with the original sound sources from which they were sampled. The experimental work with hearing impaired and cochlear implant subjects has commenced.

1 Introduction

This study investigates the perception of musical timbre by people with normal hearing, and subsequent work will compare it to that of hearing impaired subjects, and users of the multi-electrode cochlear implant developed by The University of Melbourne and Cochlear Limited. This device offers profoundly hearing-impaired individuals the ability to hear by electrically stimulating the auditory nerve fibres in the cochlea. It is now in use by more than 15 000 people worldwide, and a detailed description and a discussion of its development is provided by Clark *et al.* [1]. The system consists of two main components: the implanted receiver/stimulator and array of 22 electrodes; and an externally worn speech-processor, which includes a microphone and a transmitting coil. The intracochlear electrodes are arranged so that each one stimulates a specific region on the frequency-ordered basilar membrane. Sounds arriving at the microphone are analysed by the external processor, and then transmitted to the electrode array as a varying pattern of dynamic electrode activations. The

processor performs a frequency analysis of the incoming sound, and represents the highest amplitude spectral bands as stimuli on the corresponding electrodes. Typically, six electrodes are activated in one stimulus frame, which lasts around 4ms. The electrodes deliver pulse trains of various rates to the auditory nerve fibres. Whilst the array is permanently implanted in the patient, the external sound processor can be periodically improved and upgraded. Research continues into designing better algorithms for processing sounds.

Following the perception of speech, the appreciation of music is the next most commonly expressed requirement by users of cochlear implants. Now that these devices can offer very good performance for understanding speech, research is being extended towards improving users' ability to appreciate music. Three key aspects of musical appreciation are rhythm, pitch and timbre (or tone colour). As implants already deliver excellent temporal fidelity, and therefore rhythmic information, the task remains to improve the representation of pitch and timbre. An important dimension of timbre is spectral shape, which can be encoded in cochlear implants by the distribution of

electrode activations. Pitch can be represented by both electrode placement (given the tonotopic organisation of the basilar membrane), and by the rate of electrode pulse stimulation. These two mechanisms can be termed 'place-pitch' and 'rate-pitch', respectively. McDermott and McKay [2] have investigated musical pitch perception with cochlear implants using both of these dimensions, finding that musical intervals can be adequately conveyed in each, with place-pitch dominating when both are varied simultaneously. Rate pitch was shown to convey intervals of up to two octaves in range. Melodic recognition with implants using rate-pitch has been investigated by Pijl and Schwarz [3], who conclude that temporal stimulus cues are adequate for conveying musical pitch, at least for the lower half of the range commonly used in music. Gfeller and Lansing [4] presented research on implant users' perception of melody, rhythm and timbre, concluding that melodic and rhythmic elements of music are differentially accessible to different users. The research described below will investigate the perception of timbre, as experienced by three types of subjects: those with normal hearing, impaired hearing, and cochlear implants. This paper presents results from the first portion of the study, that dealing with the normally-hearing subjects.

Timbre, the defining characteristic which allows a listener to distinguish between two different instruments playing the same pitch at the same loudness, is a multidimensional attribute. This investigation is focussed on only one aspect of timbre - the contour of the steady-state frequency spectrum present during the sustaining portion of a note. The frequency spectrum as perceived by the listener is termed the internal spectrum. This work aims to determine how much detail is present in the internal spectra of normally-hearing listeners, and how much of this is also available to hearing aid and cochlear implant users. It will also investigate the relationship between the amount of spectral information available and subjects' ability to discriminate and identify timbres. We anticipate that people with impaired hearing and cochlear implants will not be as capable of discriminating sounds based on steady-state spectra alone, due to decreased spectral resolution. Ultimately, it may be possible to improve the design of sound processing strategies to convey more spectral information.

2 Experimental Method

Subjects were asked to undertake three tasks: a discrimination of all the pairs of stimuli from the set of timbres under investigation; the measurement of the

internal spectra of these sounds using a forward-masking paradigm, and the identification of these sounds as being derived from existing sound sources or instruments. In combination, these experiments yield a measure of the amount of spectral information available to subjects and assess its relationship to sound identification and discrimination.

Ten musical stimuli were used. These all consisted of the steady-state portions of musical sounds, with any temporal cues removed by using only short (200 ms) samples from the sustain portions of the sounds. Each stimulus had a linear rise and fall of 20 ms duration. The sounds can be categorised as follows

- Sung vowels: /a/, /e/, /i/, /ɔ/, /u/
- Acoustic instruments: oboe, bell, 'cello, organ
- Synthetic: sawtooth wave

All stimuli were presented at a pitch of B4, which has a fundamental frequency of 493.9 Hz. In addition to these musical sounds, white noise was used as a stimulus in the discrimination and identification tasks.

Each stimulus was first loudness balanced to a reference white noise which was set to a comfortable listening level of 65 dB SPL when measured with an insert earphone in an ear simulator.

For the discrimination task, each pair of loudness-balanced stimuli was presented to the subject in randomised blocks of ten trials, using a 4-interval forced-choice technique. The subject used a response box, and after pressing a button to initiate the trial, was instructed to press the numbered button indicating the one interval containing the different stimulus. A random level variation of approximately +/- 15 dB was used to avoid the use of any residual loudness cues to distinguish stimuli.

The measurement of internal spectra was accomplished using forward masking, a technique which has been applied extensively to acoustic signals, principally speech, and of which a good summary is provided by Moore [5]. This technique exploits the phenomenon whereby presentation of a sound (the 'masker') reduces sensitivity to other sounds of similar frequency for a short time (up to 200 ms) following. When a short (in this case 20 ms) pure tone 'probe' is inserted immediately following the masker, which here was the musical stimulus, an estimate can be made of the internal spectrum at the probe tone frequency, by measuring the threshold of probe audibility. In this experiment, probe tones with frequencies equal to the harmonic series based upon a fundamental frequency of 493.9 Hz were generated. These frequencies were chosen to ensure that the masked thresholds of each probe related directly to the physical spectral components present, so together they estimated the

envelope of the internal spectrum. Each probe tone was of 20ms duration, with raised cosine attack and decay portions of 5ms each, leaving a steady-state portion of 10 ms. The onset of the probe tone immediately followed the end of the masker. In this task, the subject was required to adjust the level of the probe tone until it was just audible when rapidly following the masker, using a continuously variable knob with no reference point. The masking complex stimulus was presented repeatedly with a duration of 200ms, followed by a 200ms silence before the next presentation. After every third presentation, the probe tone was inserted into the silence with no delay in the timing of the next masker stimulus. The probe could thus be judged to be audible when every third masker tone sounded different, in a 1-2-3 rhythm. When the probe was inaudible, such a rhythm disappeared. The subject was encouraged to approach the threshold level several times from both above and below, before deciding on a final value. Each spectrum was measured three times, in different sessions.

In this experiment, the masking thresholds for each

spectrum were expressed in values relative to the level of flat-spectrum masker which would produce the same degree of masking, as suggested by Moore [5]. To transform the values in this manner, the growth-of-masking functions in white noise for each probe frequency were determined for each listener. To calculate these linear functions, the masking thresholds for each probe tone frequency were determined by presenting white noise at levels of 55, 65, and 75 dB SPL, when measured as before. Gradient and intercept values were calculated for these functions using linear regression. The thresholds obtained above could then be transformed using these values.

The identification of sounds with the real instruments from which they were recorded required the subject to name each stimulus from the closed complete set of all the sound names used above, when stimuli were presented in a random sequence. Each stimulus was presented twice, following a training run in which each stimulus was presented once, with no data being recorded. No feedback was provided to the subjects concerning the validity of their responses.

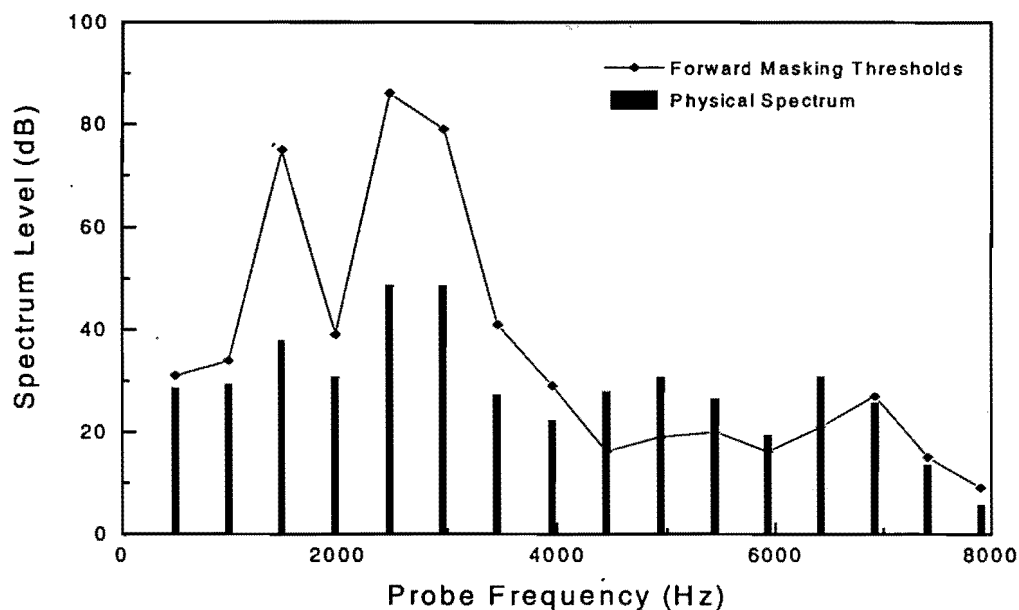


Figure 1 Physical spectrum and internal spectrum for the note B4 (fundamental frequency of 493.9 Hz) played on an oboe. The column plot shows the relative amplitudes of the frequency components in the physical spectrum for this sound, while the line plot shows the internal spectrum, measured as probe tone thresholds using a forward-masking paradigm. Whilst the Y-axis scale indicates the relative levels for both graphs, no meaningful absolute correspondence exists between the two, as only the shapes of the spectral envelopes are being compared. In this example, a good correspondence can be observed between the shape of the internal spectrum and that of the physical spectrum envelope.

3 Initial Results

So far, experimental work has been undertaken with four normally-hearing subjects. The discrimination task has yielded near-perfect results varying from 99.6% to 100% correct.

An example of an internal spectrum, obtained using the forward-masking technique and subsequent transformation, is presented in Figure 1. The column plot shows the physical frequency spectrum of the note B4 (fundamental frequency of 493.9 Hz) played on an oboe. The internal spectrum obtained by one subject using our forward masking technique is plotted as a line graph, superimposed on the column graph. A good correspondence can be observed between the envelope in the physical spectrum and the shape of the internal spectrum. This subject showed almost perfect (99.8%) discrimination of stimuli when presented in pairs, and this is probably related to his ability to accurately perceive their spectral shape.

The identification of the stimuli to the sounds from which they were sampled produced 45% correct responses for three of the subjects, while one scored 64%. The confusions made, while showing some similarity, did vary between subjects. White noise was the only sound to be identified correctly each time, while the synthesizer, /s/ and oboe were identified with reasonable consistency. The remaining sounds were identified some of the time, while the 'cello and bell were identified correctly only once each.

4 Discussion

The fact that all stimuli were distinguishable when presented in pairs was expected for the normally-hearing subjects. This portion of the experiment will yield more interesting results when applied to the hearing impaired subjects, who could be expected to show a moderate performance decrease. The implant subjects are expected to show poorer discrimination in the absence of temporal cues, because the electrode array will not be able to convey all the fine spectral detail available to the normally-hearing subjects. However, the normally-hearing subject results demonstrate that it is possible to discriminate these sounds on steady-state spectra alone.

It was interesting that subjects did not perform better on the sound source identification task, achieving only around a 50% correct recognition score. Whilst the reasonably steady sounds were adequately identified, the more characteristically transient timbres fared poorly. The bell with no striking cue, and the 'cello without its

characteristically bowed attack or vibrato, are both very unfamiliar sounds. These results illustrate the importance of temporal cues in sound source identification.

5 Conclusion

These initial results from normally-hearing subjects show the general effectiveness of the method employed in the current series of experiments on musical timbre perception. Internal spectra as measured using forward masking resemble the shapes of the physical spectra. Subjects demonstrated greater than 99% accuracy when distinguishing stimuli presented in pairs. The ability of subjects to identify steady-state spectra by name only 50% of the time, while well above the chance score of 9%, demonstrates the importance of temporal cues in sound identification.

Acknowledgements

This work was carried out with the assistance of the Bionic Ear Institute and the CRC for Cochlear Implant, Speech and Hearing Research. Thomas Stainsby is supported by an Australian Post-graduate Award. Thanks also to Laurie Cohen for informal discussions about forward masking with implantees.

References

- [1] Clark, G. M. et al., 1987, *The University of Melbourne-Nucleus Multi-Electrode Cochlear Implant*, Advances in Oto-Rhino-Laryngology, Vol. 38, Karger, Basel.
- [2] McDermott, H. J. and C. M. McKay, 1997, "Musical pitch perception with electrical stimulation of the cochlea", *J. Acoust. Soc. Am.*, Vol. 101, No.3, pp. 1622-1631.
- [3] Pijl, S. and D. W. F. Schwarz, 1995, "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes", *J. Acoust. Soc. Am.*, Vol. 98, No.2, pp. 886-895.
- [4] Gfeller K. and C. R. Lansing, 1991, "Melodic, rhythmic, and timbral perception of adult cochlear implant users", *J. Speech & Hear. Res.* Vol. 34, pp. 916-920.
- [5] Moore, B. C. J., 1995, "Frequency analysis and masking", in B. J. Moore [ed], *Hearing*, pp. 161-206, Academic Press, San Diego.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Stainsby, Thomas H.; McDermott, Hugh J.; McKay, Colette M.; Clark, Graeme M.

Title:

Preliminary results on spectral shape perception and discrimination of musical sounds by normal hearing subjects and cochlear implantees

Date:

1997

Citation:

Stainsby, T. H., McDermott, H. J., McKay, C. M., & Clark, G. M. (1997). Preliminary results on spectral shape perception and discrimination of musical sounds by normal hearing subjects and cochlear implantees. In Proceedings International Computer Music Conference 1997, Thessaloniki, Hellas, Greece.

Persistent Link:

<http://hdl.handle.net/11343/27012>

File Description:

Preliminary results on spectral shape perception and discrimination of musical sounds by normal hearing subjects and cochlear implantees