Identification of Musical Instruments by Normal-Hearing Subjects Listening Through a Cochlear Implant Simulation

Rebecca Reich[1] and Dr. Donald Eddington[2]

1 Music, Mind and Machine Group • MIT Media Lab
2 Cochlear Implant Research Lab • Massachusetts Eye and Ear Infirmary

Introduction

A cochlear implant (CI) is a neural prosthesis worn by more than 40,000 severely to profoundly hearing-impaired individuals. The primary design goal of these devices is to restore hearing sufficient for fluent communication without speech-reading. While significant progress has been made in improving speech reception, it is rare for users of these devices to enjoy music.

The goal of this research is to examine the degree to which the signal processing used by cochlear implant sound processors distorts the timbre information that musically-trained listeners normally use to identify musical instruments. Identifying the signal processing steps that eliminate and/or distort information important to the enjoyment of music is a first step towards improving implant performance by better representing music.

Background

Early studies of cochlear implant users relating to music perception involve pitch, rhythm and melodic detection and identification tasks (Eddington [1978], Townsend et al [1987], Pijl and Schwarz [1995], Gfeller et al [1997]). More recent work on timbre has assessed implant users' ability to identify musical instruments from a closed set as well as assign qualitative judgments to these signals (Gfeller [2002]). These 51 subjects achieved a mean score of 46.55% correct on eight instruments, whereas normal-hearing adults scored 90.9% correct on the same test. For implantees, confusions were distributed across all instruments, whereas the normal-hearing listeners scored more inter-family instrument confusions. A study by Fujita and Ito [1999], using 8 implantees, reported an average score of 58% recognition of five instruments. The motivation for investigating the signal processing of the implant stems from these poor results.

Methods

Simulation

MATLAB was used to implement a simulation of the Continuous Interleaved Stimulation (CIS) processing strategy, using the filter specifications of the Advanced Bionics Corporation (ABC), Clarion 1.2 implant system. Part of the code was based on work by Loizou [1999], but with several modifications including the elimination of phase information (since the implant does not perform such calculations) and the change in filter specifications as outlined by the ABC's data sheets. The following diagram illustrates the signal processing stages of the simulation:

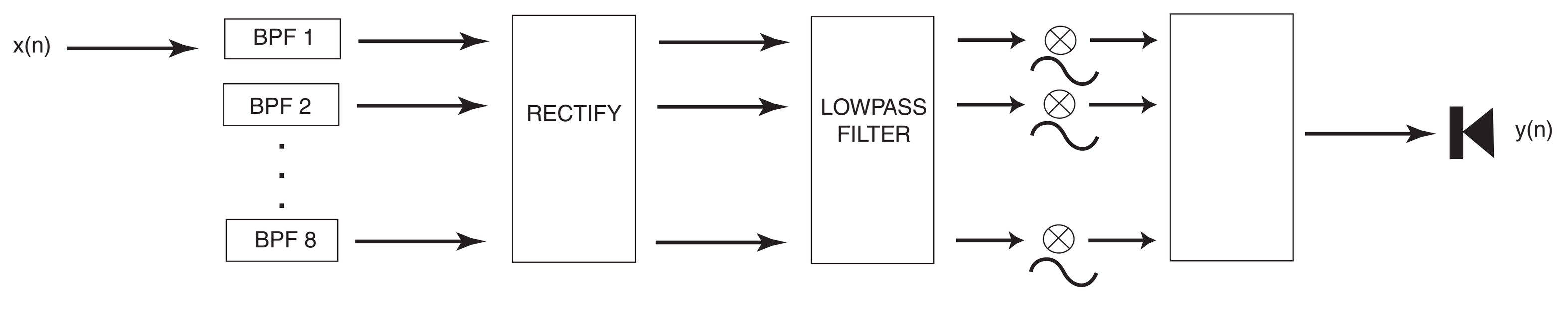

Figure 1: Signal processing stages for CI simulation

The analog input signal is digitized and downsampled to approximately 13kHz. The resulting signal, x(n), is then segmented into eight channels by a bank of bandpass filters. The center frequencies of the filters (6^{th}-order Butterworth) are logarithmically-spaced and their bandwidths span a frequency range of 350-6800Hz. The envelope of each channel is extracted by rectification(either full- or half-wave) and lowpass filtering (16-point moving-average filter). The envelope of each channel amplitude-modulates a sinusoid whose frequency corresponds to the center frequency of the channel's bandpass filter. The sum of these signals is the output, which is played over headphones to the listener.

Subjects

15 subjects took the test in two parts. Subjects were recruited by e-mail and advertisements on the MIT campus, and were compensated $10 for their participation and completion of the experiment. Subjects placed themselves in age ranges, covering a span from 17-51+ years, and reported no known hearing losses (one subject reported a possible degradation at lower frequencies due to experience playing percussion). Subjects were told beforehand that they could participate only if they were able to identify the eleven instruments used in the experiment. Subjects assessed their own musical abilities and filled out a survey on their musical background and experience.

Methods (cont'd)

Setup

Subjects were seated in a quiet recording studio and took the test on a PIII-350MHz laptop running the testing programme in Visual Basic. They wore AKG headphones and were allowed to adjust the volume to a comfortable level once before the experiment.

Experiment 1: Real (Acoustic) Instruments

Stimuli for this experiment were recordings of eight real orchestral instruments: bassoon, cello, clarinet, flute, oboe, piano, trumpet, and violin. All except the piano and violin excerpts were taken from Martin's [1999] database. The piano and violin excerpts were recorded in a studio room using ProTools™. All phrases were digitized monophonically using 16-bit resolution at 44.1kHz.

The experiment consisted of three parts:

1. Training - The training was used to familiarize the subjects with the sounds of each instrument as well as to introduce the testing software. Stimuli were short melodies played by each solo instrument (the melodies differed for each instrument). Subjects were presented with one stimulus at a time and were told which instrument was playing. They could replay the sound as many times as they wished. Subjects then selected the instrument from a list of the eight and submitted their choice, after which the next stimulus was presented.
2. Processed sounds - Different musicians for each instrument recorded two phrases of four notes. The stimuli were processed by the implant simulation, using either full- or half-wave rectification. Each stimulus was then repeated three times over the course of the experiment, and randomized in the total list, yielding 96 stimuli. Subjects would hear each stimulus once and select an instrument from the eight listed and would then submit the choice. The stimulus could not be replayed.
3. Unprocessed sounds - The stimuli consisted of the original four-note phrases, but without having been processed through the CI simulation. Thus there were 48 total stimuli. Subjects performed the same task as described in part 2 above.

Experiment 2: Synthesized Instruments

A similar experiment as described above was done using synthesized instrument signals. Eleven instrument sounds were presented: bassoon, cello, clarinet, flute, guitar, oboe, piano, alto saxophone, trombone, trumpet, and violin. Stimuli were recorded from one of three different samplers: Proteus XR, Proteus 2XR, and Roland Sound Canvas SC-55mkII. The choice of sampler for each instrument was based on the author's attempt to select the most realistic-sounding instrument samples. Stimuli were recorded into ProTools™ and were digitized monophonically using 16-bit resolution at 44.1kHz.

1. Training - The purpose of the training was the same as in Experiment 1, but here the subjects heard major scales played by each instrument as the stimuli.
2. Processed sounds - There were three different stimuli each consisting of three-note phrases, each repeated twice over the course of the experiment and randomized, yielding 132 stimuli. Subjects would hear each stimulus once and select an instrument from the eleven listed and would then submit the choice. The stimulus could not be replayed.
3. Unprocessed sounds - The stimuli consisted of the original three-note phrases, but without having been processed through the CI simulation. Thus there were 66 total stimuli. Subjects performed the same task as described in part 2 above.

Results

Real Instruments

Figure 2 (left panel) shows identification scores for the unprocessed and processed stimuli. A preliminary analysis of the data (t-test) indicates that the difference in mean scores for the unprocessed and processed cases are statistically signifi cant (0.05 level) for all instruments except for the piano and bassoon. Some instruments seem to be more affected by processing than others.

The top and bottom left panels of Figure 3 show confusions in identification for the unprocessed and processed stimuli. Confusions occurred mostly within instrument families in the case of the unprocessed sounds, whereas confusions are spread across families for the processed sounds, especially for wind instruments.

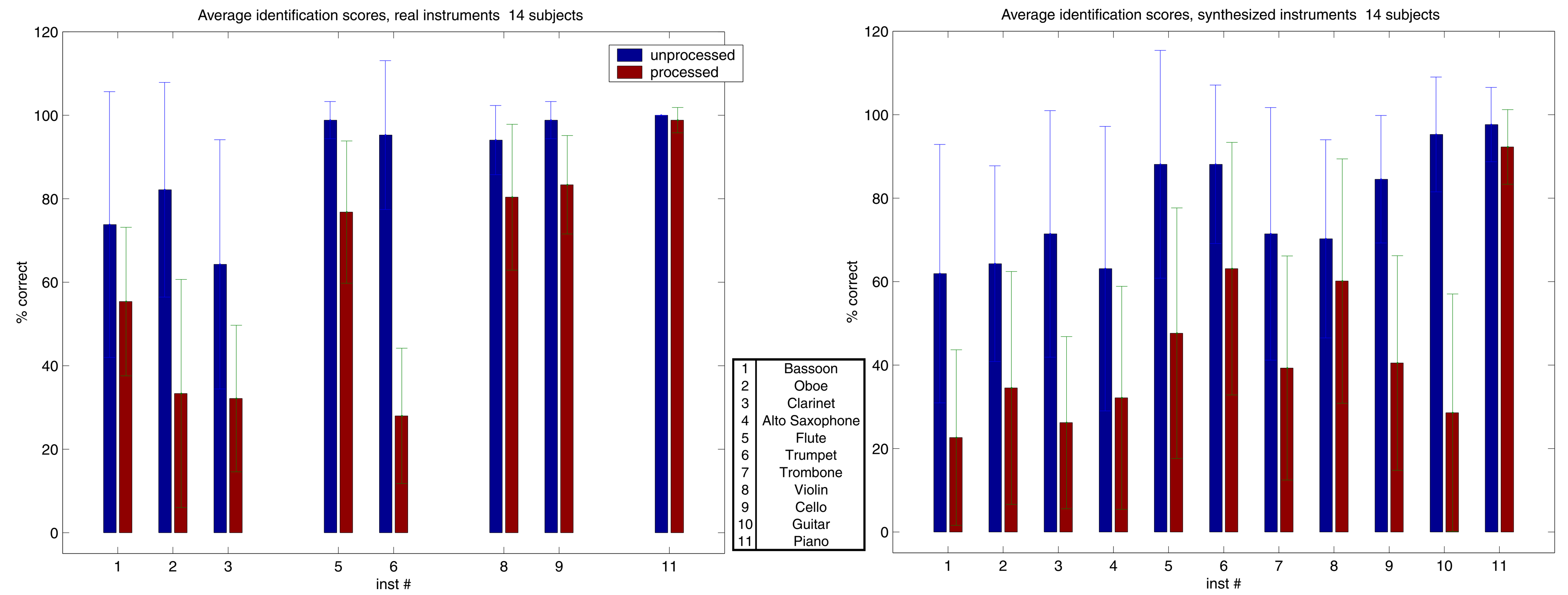

Figure 2: Identification scores using real and synthesized instruments

Results (cont'd)

Synthesized Instruments

Figure 2 (right panel) shows identification scores for the unprocessed and processed stimuli. Preliminary analysis (t-test) suggests that except for the violin and piano, the differences in means between unprocessed and processed identification scores are significant (0.05 level).

The top and bottom right panels of Figure 3 show confusions in identification for the unprocessed and processed stimuli. For unprocessed sounds, confusions occur mainly within instrument family. For processed sounds, confusions are more uniformly distributed across several instrument families.

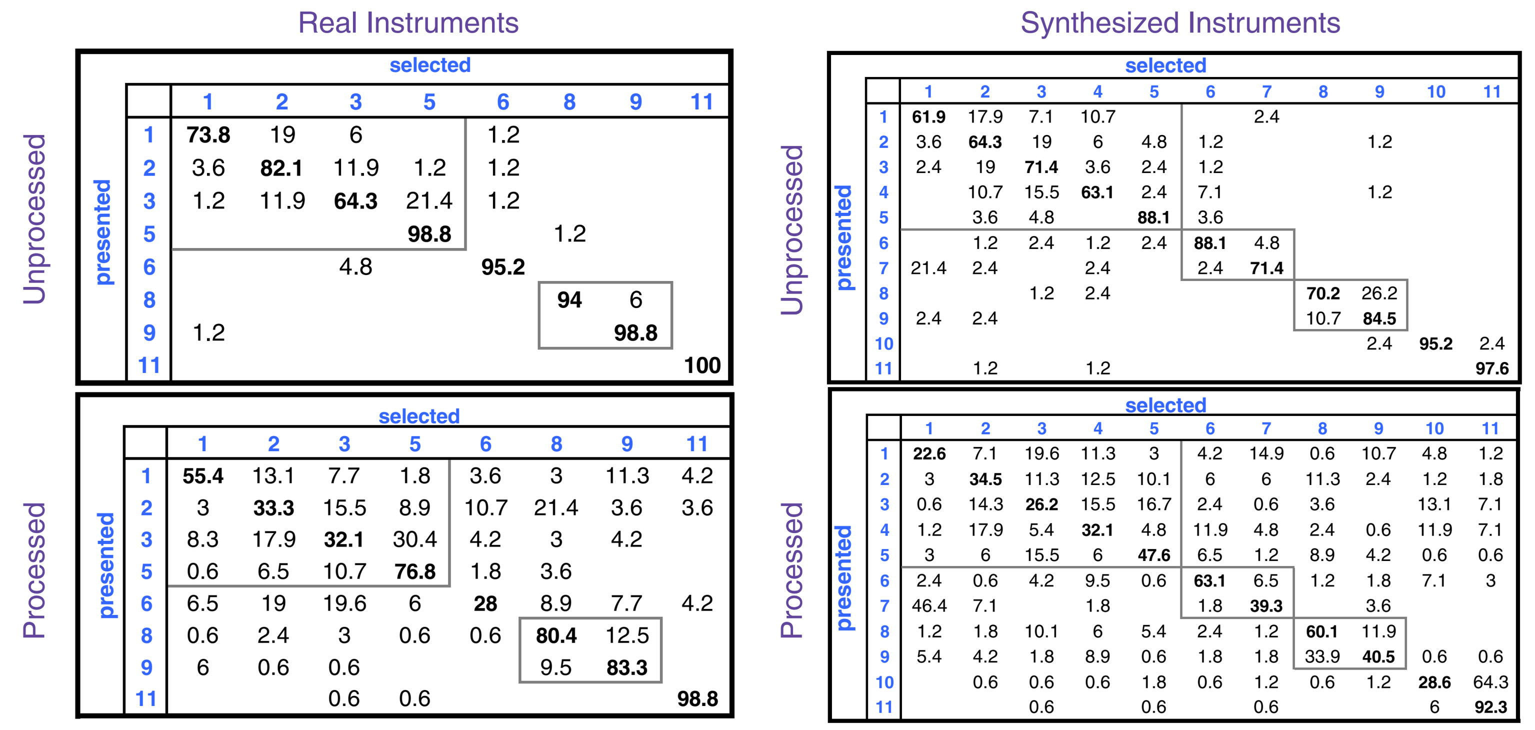

Figure 3: Confusion matrices for instrument identification

Conclusions

Simulating the signal-processing system of a cochlear implant can help normal-hearing listeners assess the ability of an implant to convey information relevant for identifying musical instruments. The decrease in identification scores from the unprocessed signals to those through the simulation indicate that the implant is not accurately representing these instruments, with the possible exception of the piano and possibly the violin. This corresponds with subjects' anectdotal comments that the piano was easy to identify through the simulation. A more detailed analysis of the data, as well as the time and frequency components of each of these instruments might indicate how to better represent them for users of a cochlear implant.

References
Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladejovsky, M.G. and Parkin, J.L. (1978). Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers. In *Trans Am Soc Art Intern Organs* 24: 1-5.
Fujita, S. and Ito, J. (1999). Ability of Nucleus cochlear implantees to recognize music. In Ann Otol Rhinol Laryngol, 108: 634-640.
Gfeller, K., Woodworth, G., Robin, D.A., Witt, S., and Knutson, J.F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. In *Ear & Hearing*, 18(3): 252-260.
Gfeller, K., Witt, S., Woodworth, G., Mehr, M.A., and Knutson, J. (2002, Apr). The effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. In *Ann Otol Rhinol Laryngol*, 111(4):349-56.
Loizou, P.C., Dorman, M., Tu, Z. (1999, Oct). On the number of channels needed to understand speech. In *J Acoust Soc Am* 106(4), Pt1: 2097-2103.
Martin, K.D. (1999). Sound-source recognition: a theory and computational model. PhD Thesis, MIT.
Pijl, S., Schwarz, D.W.F. (1995, Aug). Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. In *J Acoust Soc Am* 98(2), Pt 1 : 886-895.
Townsend, B., Cotter, N. and White, R.L. (1987, Jul). Pitch perception by cochlear implant subjects. In *J Acoust Soc Am* 82(1): 106-115.