INSTRUMENT IDENTIFICATION THROUGH A SIMULATED COCHLEAR IMPLANT PROCESSING SYSTEM

by

Rebecca Danielle Reich

B. Eng. (Electrical), Minor Arts
McGill University
Montreal, Quebec CANADA (2000)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2002

Author ____
Program in Media Arts and Sciences
August 15, 2002

Certified by ____
Barry Vercoe
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by ____
Andrew B. Lippman
Chairperson
Departmental Committee on Graduate Students

INSTRUMENT IDENTIFICATION THROUGH A SIMULATED COCHLEAR IMPLANT PROCESSING SYSTEM

by Rebecca Reich

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, on August 15, 2002, in partial fulfillment of the requirements for the degree of Master of Science in Media Arts and Sciences.

ABSTRACT

The signal processing used by cochlear implants is primarily designed to convey speech and environmental sounds, and can cause distortion of music. Although some studies of implanted listeners have demonstrated their limited ability to identify instruments, these results do not suggest whether the sound-processing system or the subjects' physiological state is the limiting factor. In this research, normal-hearing subjects determined the extent to which a simulation of the sound processing of a common cochlear implant degraded their ability to recognize instruments. The signal processing simulated in this study included bandpass filtering, rectification and lowpass filtering. Musical signals consisting of acoustic and synthetic solo instrument performances of short phrases were passed through the simulation, and presented to 25 normal-hearing participants as part of an instrument identification task. Eight acoustic and eleven synthesized common instrument timbres were used as input signals to the simulation and two rectification methods were studied for comparison purposes. Subjects were asked to identify the instrument presented, for the unaltered sounds and the sounds processed by the simulation. Identification scores for sounds heard through the simulation were significantly lower than those for unaltered sounds, due to the limited time and frequency information transmitted by the processing scheme. The results support the hypothesis that it is necessary to pursue alternative processing schemes for the implant, specifically intended for music-listening purposes.

INSTRUMENT IDENTIFICATION THROUGH A SIMULATED COCHLEAR IMPLANT PROCESSING SYSTEM

by

Rebecca Danielle Reich

Thesis Readers

Certified by ____

Dr. Donald Eddington
Principal Research Scientist
MIT Research Laboratory of Electronics

Director
Cochlear Implant Research Laboratory
Massachusetts Eye and Ear Infirmary

Associate Professor
Harvard Medical School

Certified by ____

Dr. Lorin Wilde
Research Affiliate
MIT Research Laboratory of Electronics

Lecturer
Program in Media Arts and Sciences

ACKNOWLEDGEMENTS

It would be unfair to take anything more than partial credit for this thesis, as so many have helped and contributed to make this work complete. I am indebted to Barry Vercoe for allowing me to pursue this somewhat tangential interest of mine, and for being the kind of advisor who believes in his students and encourages them to explore research topics of their choice.

I could not possibly thank Donald Eddington enough for all of his help throughout this year. I have probably caused him far more stress than he anticipated or deserved! He has taught me the value in asking the right questions, and has tirelessly guided me through the research process. He exemplifies the definition of a scientist and teacher in so many ways.

Lorin Wilde has inspired me since the first days of thesis prep class to define and narrow the scope of this project. Her advice has been so valuable in the organization and content of this thesis, and her willingness to help me at any time and her kind words are very much appreciated.

Many thanks are due to Peter Kabal, John Bradley, Jim Ryan, Dave Green and especially Meier Blostein and Mrs. Blume for helping me get to where I am in life today. My appreciation is extended to those in the scientific “audio community” who have provided advice, code or papers: Dr. Loizou, Dr. Gfeller, Keith Martin, and Joe Tierney. Neither the experiments nor this paper would been completed without Steve’s help from NecSys for all of my computer woes. Thanks to my subjects for their participation, and to Vivian Lin for her violin samples. I would also like to commend the members of the cochlear implant mailing lists who have responded so positively to my research and have inspired me to pursue this topic.

The past and present members of the Machine Listening / Music, Mind and Machine group have been so patient with my asking millions of questions. I appreciate Ricardo’s humour and advice, Wei’s help with many of my projects, Brian’s humour, help, popcorn-sharing and friendship, Nyssim’s conversations on any topic ranging from political to nutritional and her help with anything studio-related, Paris’ insights and advice and overall admirable intelligence, Joe’s and J.C.’s example of dedication and creativity, Judy’s advice and entertaining conversations, and finally Moo’s friendship and guidance throughout the two years, especially through our bi-weekly morning runs, which have truly made this experience so much richer (and healthier). Thanks to Kristie, Connie, Elizabeth and Linda for taking care of all of the nitty-gritty details. To my other friends and at the Media Lab, past and present (Stefan Agamanolis, Jon Goler, Dana Spiegel, danah boyd), it was reassuring to share this experience with others who understood so well.

My other MIT friends made my time here much more fun. Thanks to Evan, Justin and other Techiya members, and to Hillel friends for great times and laughs. I am especially grateful to Laura and Heather for their encouragement and friendship; our almost-weekly lunches and occasional skating lessons were always something I looked forward to and cherished. To the Marianopolis / McGill contingent (especially Alan, Michal, Mark, Andras, Dave), thanks for helping to keep things in perspective and for the good times. To my friends who don’t really fall into any of these categories (Mark T. & Bridget, Myles), thanks for the support. And to my Traf friends (Carrie, Julia, Joyce, Alio, Sam, Annie, Kate, Sabrina), I am so happy that we are still as close as we always were.

I cannot express my gratitude enough to my family. I would have never even known about cochlear implants had it not been for my sister Heather’s suggestion (I finally listened to your advice!). To Kevin, the Grannies, Zady, aunts, uncles, cousins and Jameson - I am lucky to have the best support network on Earth. Most importantly, to Mom and Dad - no one could ask for more love and kindness than you have provided. Your unconditional faith in my abilities surpassed my own. And finally, to Jason, who saw this process (and my emotional state) evolving on a daily basis, I am so thankful for your encouragement and love. I am so lucky to have shared it all with you.

This work was performed at the MIT Media Laboratory, and was supported by the Digital Life consortium.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

The signal processing used by cochlear implants is primarily designed to convey speech and environmental sounds, and can cause distortion of music. In this research, normal-hearing subjects determine the extent to which a simulation of the sound processing of a common cochlear implant degrades their ability to recognize instruments. We begin with the motivation for this research, an introduction to the experimental design, and an outline of the thesis.

1.1 Motivation

Cochlear implants are neural prostheses worn by more than 50,000 severely and profoundly hearing-impaired individuals. The primary design goal of these devices is to restore hearing sufficient to enable users to communicate fluently without lip-reading, and to restore an awareness of environmental sounds. Despite their improvement of speech perception, users' access to music (listening and performing) varies widely but is, in general, far poorer than normal-hearing individuals. Since music is a ubiquitous art form and is inclusive of normal-hearing participants, we address the design shortcomings of the implant to improve music perception for cochlear implant users. Our focus will be on determining the extent to which cochlear implant processing impairs instrument identification. We anticipate work similar to this will eventually lead to sound processing systems that improve the perception of music.

It is not possible to describe exactly what an implant recipient hears through these devices. There are many factors that affect perception through an implant, such as the type, duration and onset of hearing loss, the differences resulting from the surgical procedure whereby the implant is inserted into the ear, and the types of sound processing used to drive the implanted electrodes. However, it is possible to simulate the part of the implant that is worn externally and processes the incoming sound. In this way, we can enable normal-hearing listeners to assess the amount of information present in the sound-processing outputs. Using a simulation of the major signal processing elements employed by one of today's implant sound-processing systems, we will assess the ability of the processing scheme to transmit information relevant for musical instrument identification. One noteworthy caveat is that we do not claim to present sounds that are equivalent to what someone with an implant hears. Rather, we present sounds through the simulation to enable a normal-hearing person to experience the effect of the implant's processing on the information conveyed within an acoustic signal. Thus, our goal is to assess the ability of the current implant processing to transmit information relevant for musical instrument identification, using a simulation of the processor.

Instrument identification has been studied extensively in the last century to understand the perceptual mechanisms employed by humans, and more recently to enable a computer to perform the same task [Martin, 1999]. However, it is not necessary that a person be able to identify instruments in order to enjoy a piece of music. What interests us is the ability of the implant to adequately *represent* a musical instrument, given the limited amount of information that the implant transmits. It is possible to determine whether or not these instruments are being accurately represented, by having musically-trained listeners identify the instrument they hear from an audio stimulus. If musically-trained listeners cannot recognize the instruments through the simulation, then the sound processing limits the information available to cochlear implant listeners, detracting from the music-listening experience.

1.2 Research approach

We will examine the ability of musically-trained listeners to identify both real (acoustic) and synthesized instruments through the simulation of an existing implant. We wish to determine which instruments are more affected by processing than others, and to what extent. Results of an instrument identification task will be compared with previous results in the literature for normal-hearing and cochlear implanted listeners.

Once we understand how instruments are affected by processing, our goal will be to recommend design changes for future implants. More immediate applications could include a music "setting" for an existing implant, using the current parameters in an optimal configuration. We hope that the results obtained from the listening experiments will lead others to explore potential improvements.

1.3 Thesis outline

The thesis is organized into several chapters, each explained below.

Chapter 2 provides a basic overview of the anatomy and physiology of the ear and hearing loss. The history of the cochlear implant is outlined, along with its design and function. A review of simulations of implant signal processing is presented. Previous research in the area of instrument identification for normal-hearing listeners is reviewed, followed by a description of studies done with cochlear implant recipients performing various music perception tasks. First, results on tests for pitch and melody recognition are reviewed, followed by performance on rhythmic tests and then instrument identification. Finally, a discussion of surveys of implant user habits related to music listening and enjoyment is presented.

Chapter 3 outlines the methodology of our experiment. Details of the simulation, the test stimuli and testing protocols are presented. The demographics of the subject population are examined, and the effect of the processing on the audio signals is displayed visually and analyzed in the time and frequency domains.

Chapter 4 presents the results obtained from the experiments and discusses their significance.

Chapter 5 concludes the thesis by summarizing the results and suggesting areas for further research.

Appendix A contains the code for the simulation, Appendix B shows the synthesized instrument stimuli used for the experiment, and Appendix C shows the musical background survey administered to the subjects.

CHAPTER 2 BACKGROUND

This chapter provides background information and outlines previous research as context for the current research. It presents an overview of the fundamentals of hearing, hearing loss, a history and survey of cochlear implant technology and cochlear implant simulations. We examine previous experiments in musical instrument identification by normal-hearing listeners, and outline research done with cochlear implant users relating to music perception.

2.1 Physiology of hearing

Comprehensive introductions to ear physiology can be found in [Moore, 1989 or Pickles, 1988] but a brief overview will be presented here.

The peripheral auditory system acts as a transducer, converting acoustic signals into neural signals that can be interpreted by the brain. This system is comprised of three parts: the outer, middle and inner ears. Figure 1 shows these three parts as one system in detail.

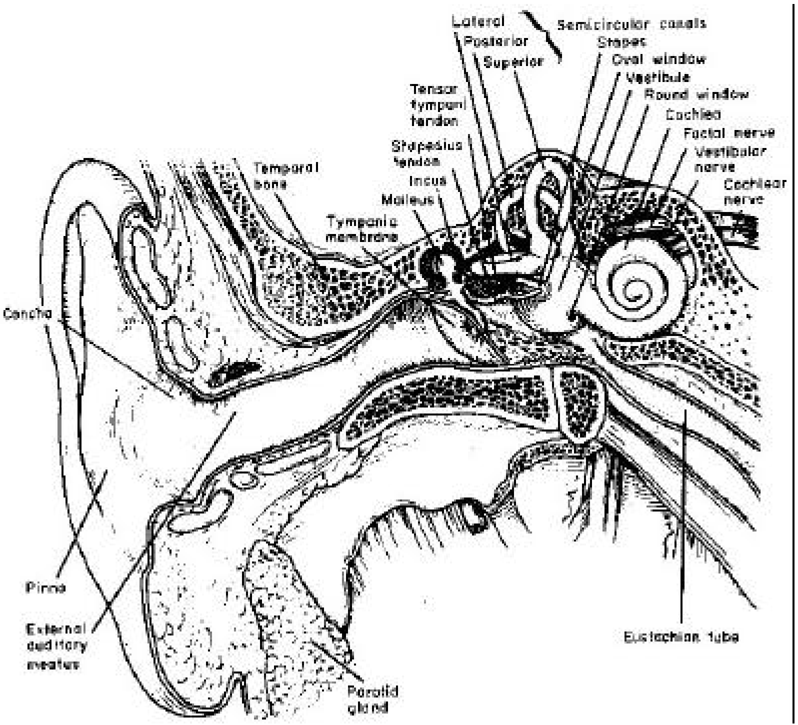


Figure 1: Anatomy of the ear [Kessel and Kardon, 1979]

The outer ear consists of the pinna and the auditory meatus (canal), which act as resonant filters, increasing the transmission efficiency of incoming signals in the 2-7kHz range. The outer ear is also responsible for aiding in sound localization as it modifies high frequencies of the incoming sound. The sound vibrations travel down the canal and act on the tympanic membrane (eardrum).

The middle ear consists of three small bones (collectively called ossicles): malleus, incus and stapes (commonly known as hammer, anvil and stirrup). The bones mechanically conduct the sound vibrations from the tympanic membrane to the oval window of the cochlea in the middle ear.

The cochlea is a snail-shaped structure filled with nearly incompressible fluids, located in the inner ear. A diagram of the cochlear appears in Figure 2.

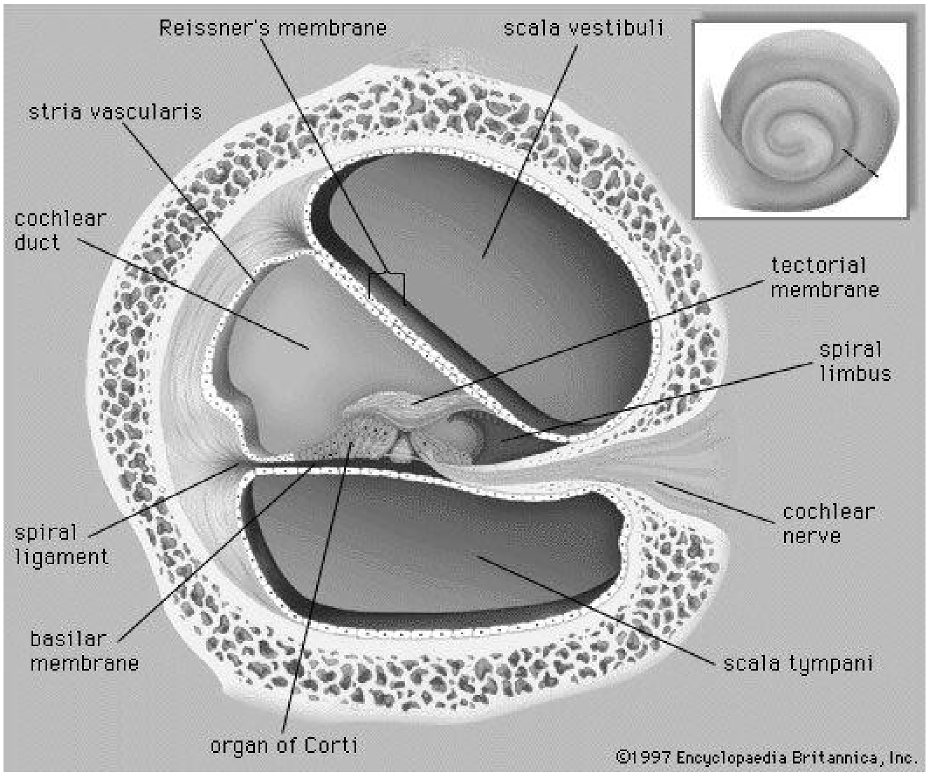


Figure 2: Diagram of the cochlea [(1997) *The Cochlea*]

The basilar and Reisner's membranes run along its length, partitioning the cochlea into three compartments. The basilar membrane is mechanically tuned and plays an important role in distributing sound energy by frequency along the cochlea's length. High-frequency energy vibrates the basilar membrane near the oval window (basal) end of the cochlea. As one moves from the base to the apex, the basilar membrane motion represents energy at progressively lower frequencies, due to its decreased stiffness. The term "tonotopic organization" of the basilar membrane is often used to describe this frequency-mapping mechanism. Figure 3 shows the peaks of amplitude of excitation of the membrane as a function of frequency along the length of the cochlea.

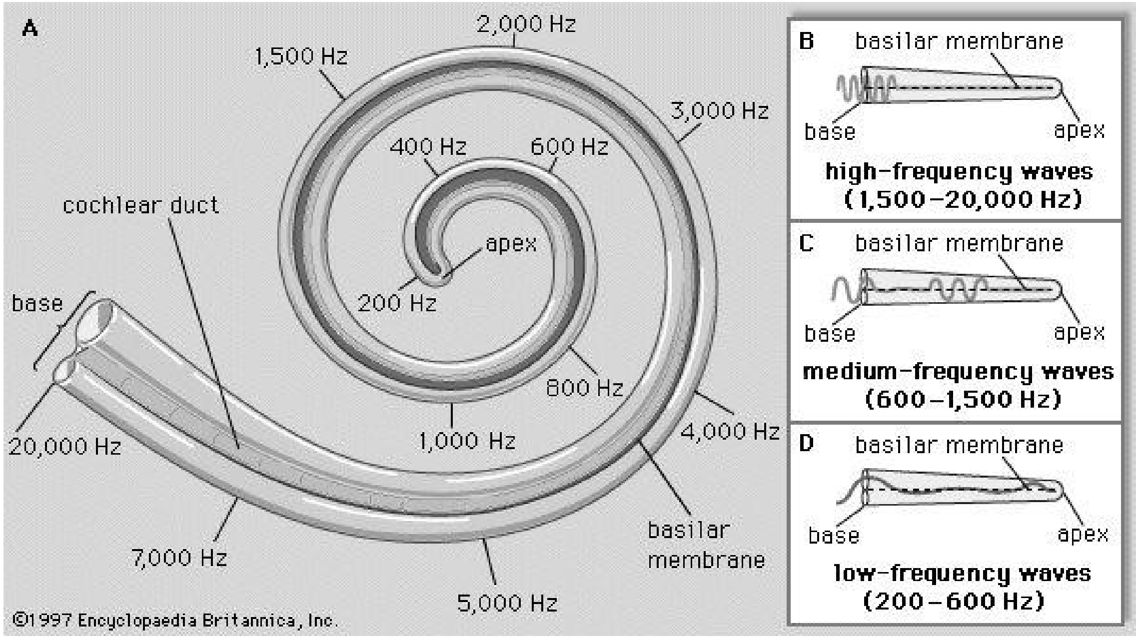


Figure 3: Tuning of the basilar membrane [(1997) *Basilar Membrane*]

Measurements of vibration patterns along the membrane show that the responses are not linear: an increase in amplitude of the incoming sound will not cause the response to grow in direct proportion, over the dynamic range of hearing.

The process of transduction of basilar-membrane vibrations to electric spikes on the auditory nerve is performed by hair cells that lie adjacent to the membrane. There are approximately 3500 inner hair cells that control the spike activity on approximately 30,000 auditory nerve fibers.

Two types of cues have been suggested as important in coding pitch. In the case of the “place cue”, a tone’s pitch is attributed to the stronger response of nerve fibers enervating hair cells at the place on the basilar membrane responding maximally to a given frequency. However, one can also examine the patterns of nerve firings (that is, intervals between spikes) that occur for a tone of a given frequency. This “temporal cue” is thought to be important for tones below 3kHz where spike intervals carry information about a tone’s pitch. In general, it is unknown how these two types of information are combined in the formation of pitch sensation [Pickles, 1988].

In summary, the frequency of the incoming sound is resolved spatially on the basilar membrane, with maximal membrane displacement producing more nerve firings in the fibers enervating hair cells located at the place on the membrane corresponding to that frequency. The response magnitude (firing rate) of an auditory nerve fiber shows a bandpass characteristic when plotted as a function of frequency.

2.2 Hearing loss

Hearing impairment affects more than 22 million Americans, of which possibly 738,000 are cases of severe to profound hearing loss (a loss of >70dB in auditory sensitivity), which cannot be restored by conventional hearing aids [Blanchfield et al., 1999 and Tierney et al., 1994]. A brief discussion of the types of hearing loss follows, to provide motivation for the invention of the cochlear implant.

Given that several mechanisms exist which contribute to the hearing process, there is potential for damage at several points within the ear. The two main types of hearing loss are classified as either *conductive* or *sensorineural*. Damage to the region lying exterior to the cochlea causes conductive loss, while damage either within the cochlea or auditory nerve is termed sensorineural loss. Conductive loss, a mechanical loss, can often be treated with medical interventions or remedied with hearing aids, since the loss is usually in the form of attenuation across frequencies. The other form of loss, sensorineural loss, is more likely due to problems in the cochlea (which can arise from a variety of origins) rather than with the auditory nerve. A common form of loss due to aging is called presbyacusis [Pickles, 1988], whereby sensitivity is gradually lost, especially at higher frequencies. Some causes are present from birth while some can have a more sudden onset, such as acoustic trauma, or side-effects from medications. Often it is the hair cells that are damaged, which cannot be regrown or replaced. Hearing aids may not provide adequate when frequency resolution cannot be restored by simple amplification, and clarity is compromised due to the limitations of the devices. Cochlear damage can lead to decreased sensitivity, first appearing at higher frequencies (for hearing losses originating from prolonged exposure to noise, or presbyacusis). The spatial selectivity of basilar-membrane displacement (i.e., the bandwidth of the filter centered at a particular frequency) also broadens with sensory impairment, causing a decreased sensitivity to differences in frequencies of tones. Another phenomenon associated with hearing loss is loudness recruitment. This term refers to the phenomenon whereby as the level of a tone is increased, the sensation of loudness increases in an unusually rapid manner. The result is a reduced dynamic range, and sounds that are audible might not be easily intelligible. These forms of sensorineural loss, which cannot be corrected with hearing aids, prompted the investigation of direct electrical stimulation of the auditory nerve, which is the basis of how the cochlear implant came to be. Candidacy for an implant is different for adults and children. For adults, sensorineural loss must be bilateral (in both ears) and severe to profound (>70dB at 500Hz, 1000Hz and 2000Hz). Children aged 18 months to 17 years old must have profound, bilateral loss (>90dB at 1000Hz). For younger children, they must demonstrate slow progress in development of auditory skills while participating in aural rehabilitation, and older children need to have tried hearing aids for three to six months [FDA, 2001].

2.3 A brief history of cochlear implants

Electrical stimulation of the auditory system can be dated back to the 18th century. Alessandro Volta was one of the first investigators of electrical stimulation, eliciting hearing sensations by applying a voltage gradient across his ears [House, 2002]. Only a few others experimented with electrical stimulation for the next half-century. By the 1960’s, groups led by Simmons, House and Michelson implanted human subjects with electrode systems placed in and around the cochlea (see House [House, 1976] for a comprehensive timeline with more detail of implanted subjects). In addition, the advent of technologies such as pacemakers was encouraging for bio-compatible devices such as cochlear implants. The first wearable sound processing systems included a single analysis channel driving a single intracochlear electrode. While these devices significantly improved lip-reading and conveyed environmental sounds, they did not provide sufficient information for patients to understand speech without lip-reading. In the 1970s, Eddington [Eddington, 1978] reported results for one patient with an eight-electrode (six intracochlear) implant. Other researchers (e.g., Chouard, Michelson and Merzenich [Michelson et al., 1975]) began working with multi-channel sound processors in the early 1980s. Early devices were large, consumed large amounts of power and their modest benefits restricted their use to post-lingually deafened (onset of deafness after language has been acquired) adults. Since the 1980s there have been several iterations of cochlear implant design. On average, modern cochlear implants provide much better speech intelligibility compared to the first devices and in some cases enable patients to converse fluently without lip-reading.

2.4 The cochlear implant - design and function

A cochlear implant consists of several parts. A microphone worn at ear level captures the incoming sound. The sound is digitized and analyzed by a processor, originally worn as a pager-sized box and now miniaturized to be worn behind the ear

(BTE). The processor divides the signal into several channels based on frequency, translates the information in each channel into instructions that are transmitted to and control an implanted receiver / stimulator that drives the implanted electrode array. The array of electrodes consists of six to 22 intracochlear electrodes distributed along the length of the cochlea. Stimuli delivered to an electrode preferentially excite the nerve fibers nearby. A diagram of an implant system is shown in Figure 4.

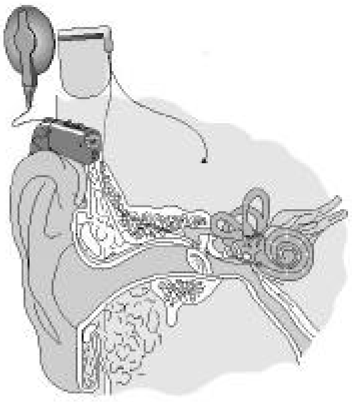


Figure 4: Illustration of cochlear implant. The processor worn behind the ear sends a signal to the transmitter that includes the stimulation instructions and power to operate the implant. A receiver/stimulator implanted under the skin interprets the instructions and delivers the stimuli [(2002) *Med-El*]

It is the signal processing system that is of most interest for the research of this thesis. Over the years, different strategies have been designed for converting the sound to stimuli delivered to the electrodes (see Loizou [Loizou et al., 1999] for a comprehensive history of the signal processing of the implant).

One common processing strategy is outlined in Figure 5.

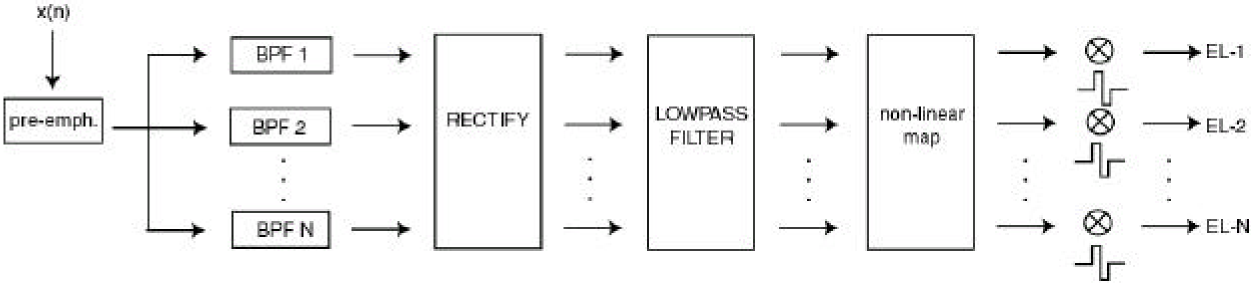


Figure 5: Signal processing stages for a continuous interleaved sampling (CIS) sound processing strategy. Input `x(n)` is bandpass-filtered, and each output is rectified, lowpass-filtered, and mapped to the listener's dynamic range. These signals amplitude-modulate pulse trains that are each presented to an electrode.

In this system, the signal is pre-emphasized to boost the higher frequency components. The signal is then divided into channels by a set of bandpass filters and the envelope from each channel is extracted. The envelopes are compressed logarithmically to approximate abnormal growth of loudness over the relatively small dynamic range of electric stimulation. The compressed envelopes modulate the amplitude of a train of biphasic pulses (short pulses with a brief positive and negative component, making the DC value zero) that are delivered to that channel's electrode. Because the inter-electrode spacing is often less than 1mm, the interaction of adjacent electrical fields during stimulation can result in significant distortion. The strategy shown in the figure addresses this issue by interleaving pulses consecutively in time across electrodes, avoiding simultaneous stimulation

[Wilson et al., 1991]. The pulse repetition rate can be customized for each listener, but then remains fixed. More detail about this processing will be seen in Chapter 3.

2.5 Cochlear implant simulations

From the foregoing description of the processing done by a cochlear implant, it is clear that a significant amount of information is discarded before sound stimuli are presented to an implant recipient's nervous system as electric signals. For instance, the resolution for representing the instantaneous spectral shape is limited by the number of analysis channels in use (typically six or eight at any given instant). The envelope extraction stage of each channel also discards any information in the fine time structure of each channel's signal.

Several investigators have adapted cochlear implant sound processors to provide acoustic outputs for normal-hearing listeners, in order to get a sense of the impact that this processing has on information required for speech reception. For normal-hearing listeners, one must mimic the effect of presenting an amplitude envelope to a particular location in the cochlea by stimulation of an electrode. This can be accomplished by using modulated bands of noise or sinusoid tones to place the low-bandwidth information of the envelope in the appropriate cochlear position. Using these methods in simulations of cochlear implant sound processing, investigators have compared speech perception of normal-hearing subjects listening to unaltered (natural) and processed speech.

A detailed review of earlier simulations can be found in Weis [Weis, 2000]. A recent simulation conducted by Shannon et al. [Shannon et al., 1995] focused on the amount of spectral and temporal information needed to recognize speech. Their simulation divided the signal into four frequency bands and the envelope in each was extracted by half-wave rectification and lowpass filtering (various cutoffs were implemented and tested). The envelopes modulated bandpass-filtered noise (the noise was filtered by the same bandpass filters in the analysis stage of the processing) and all bands were summed and presented over headphones. The stimuli consisted of medial consonants (a consonant between two vowels), vowels and simple sentences. After training on all sounds, eight listeners were asked to identify each stimulus (closed-set). Variation of the lowpass cutoff frequency did not result in significant performance differences, and it was determined that even with only low-bandwidth envelope information, high speech recognition scores (90% correct word identification) could be achieved with three analysis channels.

Dorman et al. [Dorman et al., 1997] performed a similar simulation but investigated sine tones as well as band-limited noise as carriers. Signal pre-emphasis was added and the number of channels was varied between 2-9. As the number of analysis channels increased, vowel recognition scores reached an asymptote at eight channels while for sentences, only five channels were required. There was only a small difference found between using sinusoidal versus noise band carriers.

Loizou [Loizou et al., 1999] used a simulation employing sinusoidal carriers to examine the number of channels required to understand speech with multiple talkers. Stimuli consisted of phonetically-rich sentences (the TIMIT database). Five channels were required to achieve a 90% correct score and beyond eight channels, the results were asymptotic. This simulation included phase information not included in cochlear implant processing.

The most detailed simulation has been done by Weis [Weis, 2000], who not only simulated the signal processing of the cochlear implant, but also some perceptual limitations associated with hearing impairment. His simulation was used to test subjects' scores on speech recognition tasks, and results were similar to those conducted on implanted listeners.

In summary, simulations can give researchers insight into the amount of information being passed through a cochlear implant processing system. We will use simulations similar to those described above to investigate the degree to which the signal processing of a cochlear implant limits the information that musically-experienced listeners use to identify instruments.

2.6 Previous research

We will now examine prior studies in several music-related areas to provide context for the current research. First, experiments in musical instrument identification by normal-hearing listeners will be reviewed, and features that are useful for identification will be highlighted. Following this will be a review of research in cochlear implant recipients' abilities to discriminate or identify pitch and melody, perceive rhythm information and identify musical instruments. These results motivated the experiments that will be described in later chapters.

2.6.1 Instrument identification - normal hearing

Before we examine the effect of cochlear implant processing on instrument identification, it is useful to have some knowledge of the performance of normal-hearing listeners identifying instruments, and the features commonly used to perform this task. Often, the term "timbre" is used to describe the quality of a sound produced by a musical instrument. A formal definition states that timbre is "that attribute of a tone by which a listener can judge that two sounds of the same loudness and pitch are dissimilar" (ANSI 1973). Timbre can be a function of several signal parameters, including the attack and decay transients and the frequency spectrum. Helmholtz [Helmholtz, 1954] showed that the relative amplitudes of harmonic partials that form a periodic tone primarily determine the tone's sound quality, regardless of the phase components. The *formant theory* of musical quality states that the quality of an instrument's sound arises from the "relative strengthening of whatever partial lies within a fixed or relatively fixed region of the musical scale" [Bartholomew, 1942]. In this theory, it is the *changes* in the spectrum that influence the musical quality of the sound. The strengthening of the partials is due to the resonances produced by the body of the instrument. Timbre is thought by some to be a meaningless term [Martin, 1999], but it is important to be aware of how people have used this term when reading the literature.

It is also important to be aware of the term "identification", which can mean different things depending on the experiment design. For example, subjects can be asked to freely write down the instrument that they believe they are hearing (open-set). Alternatively, they can select an instrument from a list that includes some labels of instruments not heard in the experiment (closed-set with distractors). Finally, subjects can be asked to choose from a list of instruments that only contains those heard in the experiment (closed-set). We will refer to all of these tasks broadly as instrument identification.

Early experiments in timbre identification used single-notes as stimuli. Eagleson and Eagleson [Eagleson and Eagleson, 1947] tested 14 musicians and 13 non-musicians in two open-set testing situations: one where listeners heard only the sustained portion (no attack) of a single tone over a public-address system (PA), and one where listeners heard the entire tone live. In general, the average number of correct scores for instrument identification was lower for sounds over the PA (average around 35% for sounds over the PA, 50% for direct sounds), but the significance of these results are unknown. The cymbals, violin, trumpet and bells were considered easier to identify than the alto horn, piccolo, flute, clarinet and saxophone. The piccolo and alto horn were the most difficult to identify. It is important to note that the attack transient seemed to be important for identification.

Saldanha and Corso [Saldanha and Corso, 1964] studied identification of single, sustained tones for 10 instruments (anechoically recorded, with and without vibrato), by 20 musically-trained subjects. The tones were spliced to test various initial and final transients in combination with the steady-state portion of the tone for each instrument. The test was closed-set with 39 possibilities and 10 instrument stimuli. Highest identification scores occurred for wind instruments, when the initial attack and short steady-state portion of the tone were present. Poor identification scores occurred when the steady-state portion was played alone, or in combination with the final portion of the tone. Vibrato improved identification scores, and scores also improved with practice. In this study, the clarinet, oboe and flute were identified correctly more often than violin, cello and bassoon. Overall the mean correct score was 40%.

In Berger's [Berger 1964] examination of wind instruments (flute, oboe, clarinet, tenor and alto saxophone, trumpet, cornet, French horn, baritone and trombone) playing sustained single tones (F4, 349Hz), several conditions were examined: 1) unaltered notes played for a total of five seconds, 2) sustained portion only (first and last .5s removed), 3) tone played in reverse, and 4) tone passed through a lowpass filter (keeping only the fundamental). Thirty band students identified instruments from a known set. In the unaltered case, university band players scored a mean of 59% correct for all instruments. Tones played backward resulted in an average score of 42% correct. A 35% correct mean score was achieved for tones with the attack and decay portions removed. When the upper harmonics of the tone were eliminated by filtering, the mean score was 18%. For all conditions except filtering, the oboe was easiest to identify, while the flute and trumpet were the most difficult. Confusions occurred mostly within the instrument category (e.g., woodwinds and brasses).

Strong and Clark [Strong and Clark, 1967] investigated identification of synthesized woodwind tones while altering spectral and temporal information. Eight musically-literate subjects identified the perturbed tones of the trumpet, trombone, tuba, French horn, oboe, English horn, bassoon, flute and clarinet. Results showed that, in general, higher identification scores occur if the temporal and spectral envelopes from an actual instrument are used simultaneously (that is, instead of "cross-breeding" instruments by swapping spectral and temporal information). Perturbing the tones of the instruments led to various conclusions regarding the important features of each. The results suggest that spectral envelope is more important than the temporal structure for the oboe, clarinet, bassoon, and tuba. The importance of spectral and temporal information is similar for the trumpet, trombone and French horn. The temporal envelope of the flute is more important than the spectral envelope. Interfamily confusions were lower for cases where the spectral envelope contributed more to correct identification.

Grey's study [Grey, 1977] examined the perceptual relationships between 16 instruments using computer-generated single tones and 20 musically-sophisticated listeners. Subjects were told to rate the similarity of pairs of stimuli. Multi-dimensional scaling yielded a three-dimensional solution in space to best represent the perceptual similarities. These three dimensions are the following: 1) spectral energy distribution, 2) synchronicity of the attacks and decays of the upper harmonics, and 3) presence of low-amplitude, high-frequency energy in the attack. Another experiment tested 22 musically-trained listeners' instrument identification abilities with single-tone stimuli. Subjects received feedback during the first set of trials and thus had the opportunity to learn the instruments. The number of correct scores improved with practice, from 60% on average in the first session to 84% by the fifth session.

Kendall [Kendall, 1986] questioned the notion of using single tones as stimuli for instrument identification, based on previous research involving note transitions. He was also one of the first to use a true forced-choice paradigm for testing. Musicians recorded single tones and three legato phrases from folk songs on the clarinet, violin and trumpet. Stimuli were edited for various testing conditions: normal, time-variant steady-state alone, transients alone, and static steady-state (with and without transients). A test tone or phrase was played by one instrument, followed by three "choice" tones or phrases played by three instruments (one of the instruments was the same as in the test phrase, played by a different performer). Subjects (college music and non-music majors, numbers of each are not specified) had to choose one of three "choice" tones/phrases that sounded most similar to the test tone/phrase. The responses to the transients alone and steady-state signals (single tones) were similar to those obtained by Saldanha and Corso (43%). For single notes (normal signals), mean scores of 50% and 58.3% correct for non-music and music majors respectively were obtained. Scores for unedited tones were similar to those using only transients. These results agree with earlier studies in finding that transients alone were sufficient (but, in this case, not necessary) for identification. For full phrases, 74.2% and 94.6% were correctly matched by the non-music and music majors, respectively. Transients were neither sufficient nor necessary to categorize the three instruments in the phrase context. This is the first evidence to contradict the studies on single notes. Overall, the experiment showed that using whole phrases yielded higher scores than the single notes.

Handel (Handel, 1993) reviews research in timbre identification, and states that the energy spectrum of the harmonics of a tone in part determines timbre and identification. The steady-state portion of a tone is sufficient for identification of instruments, but produces roughly 2/3 of the performance when using the whole signal. The initial transient and attack also partly determine timbre. According to Handel, it is unclear whether or not initial transients alone can enable a high level of instrument identification.

More recent work in instrument identification by humans and computers has been done by Martin [Martin and Kim, 1998 and Martin, 1999]. Some of the features used for computer identification included pitch, spectral envelope, intensity, spectral centroid (the “balancing point” of the spectrum, often associated with perceived “brightness” of a sound), onset characteristics of harmonics, and vibrato and tremolo strength. When testing computer identification, vibrato strength and onsets were important features for almost all families of instruments. Human identification abilities were tested using isolated tones and phrases played for 14 subjects with substantial musical background. The stimulus could be replayed as often as the subject desired, and the list of possible choices was 27 (though several of those were not in the test set). For single tones, the mean score was 45.9% (91.7% correct for instrument family). For 10-sec phrases, subjects scored 66.9% correct (96.9% correct for instrument family). Confusions occurred mainly within family (e.g., violin and viola confusions occurred particularly often).

In summary, it seems that musical phrases make instrument identification easier than single tones. Average scores for identification range from 45%-95% for non-perturbed sounds, with parameters influencing performance including the details of the subjects’ task, their musical training, the type of stimulus (single-tone versus phrase) and the characteristics of the stimuli (e.g., attack transients, steady-state duration, vibrato and spectral energy distribution).

2.6.2 Cochlear implants and music perception

Subjective tests on implant recipients have focused mainly on the ability to recognize speech (vowels, consonants, words and sentences), with and without the aid of lip-reading. Tests evaluating the benefits provided by cochlear implants for identification tasks related to music perception are few, though this area has been gaining some interest recently. This section will highlight and summarize relevant research.

2.6.2.1 Pitch and melody

Early work examining pitch perception by implant recipients using direct stimulation was done by Eddington [Eddington, 1978]. One post-lingually deafened subject with a 6-electrode implant was studied. Both the place and frequency of stimulation were varied. When the place of stimulation was varied, the subject ordered the relative pitch across electrodes consistent with the tonotopic organization of the basilar membrane (later research by Busby et al. [Busby et al., 1994] also reported consistency with the tonotopic arrangement when testing nine postlingual subjects with 22-electrode implants). When the rate of stimulation was varied from 60-400Hz, the subject studied by Eddington reported perception of a rising pitch. When 100,175, and 300Hz stimuli were each presented to each of the six implanted electrodes, the subject was asked to place them on a scale from 0-100 in terms of pitch. The perceived pitch increased with the higher electrode numbers and with higher stimulation rates. Testing of melody recognition used five tunes of equal note duration (using computer-generated stimuli with varied rates to elicit a change in pitch) on single electrodes. The subject recognized three of five melodies played on electrode 3 but could not recognize any tunes played on other electrodes. The subject could correctly identify relative pitch changes on any of the electrodes.

Research by Townsend et al. [Townsend et al., 1987] examined pitch perception of three post-lingually deafened subjects (two with eight-electrode implants and one with six) using direct electrical stimulation. In examining the variation due to place of stimulation, one subject had clear tonotopic results, another less so, and the third hardly noticed differences in pitch between electrodes. Simultaneous stimulation of two electrodes produced some intermediate tones, without producing a perception of two distinct sounds. These results indicated that a continuous range of perceived pitches might be generated (which is an interesting concept that could be useful for better melody perception). With varying rate of stimulation, subjects differed in their ability to discriminate pitch. In another test, it was found that increasing the amplitude level produced small increases in perceived pitch for all subjects, in a manner consistent with data for normal-hearing subjects [Cohen, 1961].

Pijl and Schwarz [Pijl and Schwarz, 1995] studied 17 postlingually-deafened subjects using the Nucleus implant. The first experiment tested melody recognition of rhythmically-intact common tunes using direct electrical stimulation. Subjects were asked to write either the title, a line from the lyrics or “unfamiliar” for each melody (open-set with 30 tunes). When pulse rates were varied, pilot studies for rates in the 100-500 pulses-per-second (pps) range were reported as sounding more musical. For a 100 pps stimulation rate, results showed 11 of 17 subjects scored better than 40% correct. Higher scores were reported for subjects who had played a musical instrument. A second test presented melodies (closed-set) without rhythmic cues to three subjects with limited to no musical training. Melodies were presented on three electrodes at once, all on basal, mid-portion or apical regions of the array. Six different pulse rates (seven for one subject) were tested in combination with the three sets of electrode positions. The best performance occurred for low pulse rates at apical electrodes (which were also reported as sounding more musical and pleasant). A third test involved interval identification involving three subjects. They had to indicate if the second note in an interval was in tune, too low or too high relative to their memory of the interval size for that melody. There were five presentations of 12 intervals (integer relations between two tones of different pulse rates) on one apical electrode. The musical intervals defined by the subjects corresponded to the same ratios between tones as intervals between acoustical frequencies perceived by normal-hearing, non-musicians.

Gfeller has reported a series of related studies on implant recipients and music perception. In one study [Gfeller and Lansing, 1991], 18 post-lingually deafened subjects were evaluated for their ability to distinguish between pairs of melodic patterns (two to five notes) with identical rhythmic patterns. Melodies were played over headphones with subjects using their commercial sound processors. A mean score of 78% correct was achieved. In a later study [Gfeller and Lansing, 1992] involving 34 subjects performing the same task, the mean score was 77.5% correct.

Research by Gfeller in 1997 [Gfeller et al., 1997] included 17 adult implant recipients (it is not mentioned if subjects were post-lingually deafened) and 35 normal-hearing participants, and examined the difference between two sound processing strategies in implants using 22 electrodes. One strategy, F0F1F2, extracts the fundamental frequency and estimates the first two formants. The MPEAK strategy uses additional high frequency spectral energy in its analysis, potentially enriching the quality of the sound. The same standardized test as in the 1991 study was used, testing perception of isolated sequential pitch (identifying similarity in pairs of tones) and rhythmic patterns. In this case, the stimuli were computer-generated square-wave tones, played over loudspeakers. Average scores for the melodic task ranged between 70-80% and no difference between processing strategies was noted over time. Normal-hearing listeners perform significantly better for the same task.

McDermott and McKay [McDermott and McKay, 1997] examined pitch perception in one implant recipient who, before the onset of hearing loss, was an instrument tuner and thus had prior knowledge of music to judge musical intervals reliably (it is not mentioned if he was post-lingually deafened). The two tasks were interval estimation and interval production by adjustment. Stimuli were delivered to one electrode directly (three electrodes were tested in all), and the stimulus frequency was determined by varying the pulse repetition rate, electrode position or some combination of both parameters. The subject was asked to indicate which of two stimuli had a higher pitch. Overall it was determined that an interval as small as a semitone (corresponding to a 5.9% ratio in frequency) could be distinguished, within a range of two octaves. Another experiment asked the subject to adjust the perceived pitch of one stimulus to match a reference. All intervals produced were close to or larger than the target interval. A third test asked the subject to name the musical interval between two tones as rates were varied. Increasing electrode separation elicited responses corresponding to larger intervals. Responses to smaller intervals were more accurate than for larger intervals.

Fujita and Ito [Fujita and Ito, 1999] evaluated the ability of eight post-lingually-deafened adults with implants on their ability to recognize nursery melodies. Twenty songs were used, which were familiar to most people. Ten were played from a cassette and the others played by synthesizers (no vocals). When subjects indicated their choice of song in an open-set identification task, they scored an average of 39% correct. In a closed-set experiment with 10 other tunes (with which subjects indicated they were familiar), subjects scored an average of 53% correct. In a test with four songs having the same rhythm and similar pitch range, played on a synthesizer, subjects did not score better than chance (25%) on identification. However, in this experiment, the researchers played the melodies on the keyboard for each trial, claiming “great care [was taken] to play each song at the same tempo and rhythm”, which may not have produced consistent stimuli. For interval identification, subjects heard intervals ranging from two to 12 semitones and were required to indicate which pitch was higher. Results were bimodal: those who could distinguish intervals of four to 10 semitones apart, and those who could not distinguish intervals of 12 semitones.

2.6.2.2 Rhythm

Perception of rhythm by recipients of cochlear implants tends to be more accurate than perception of pitch. In Gfeller’s 1991 study, performance on the rhythmic test (similar to the melodic test, where pairs of rhythmic phrases are played and subjects judge whether the pair is the same or different) achieved a mean score of 88% correct. Gfeller’s 1992 study confirmed this with a score of 85%. Gfeller’s 1997 study used a subset of the 1992 rhythmic patterns, evaluating the two processing strategies mentioned earlier (F0F1F2 and MPEAK). One test asked subjects to identify where, in a sequence of six pulses, a shorter pulse was inserted. Normal-hearing subjects scored significantly higher for this task compared with implant recipients. Another test involved identifying similarity between pairs of rhythmic stimuli. Scores ranged from 70-85% correct for this task. Normal-hearing listeners’ scores were not significantly different than those of the implant recipients. No significant differences between strategies were noted for either rhythmic test.

2.6.2.3 Instrument identification

Gfeller’s 1991 study also examined implant recipients’ qualitative assessments of musical instrument sounds. Subjects heard short, recorded excerpts of solo melodies played by the violin, cello, flute, clarinet, saxophone, oboe, bassoon, trumpet and trombone played over headphones. Subjects rated the instrument “quality” with bipolar adjectives (such as “beautiful/ugly”) and were asked to identify the melody and instrument. Generally familiar melodies were used. The percentage of subjects who rated each instrument as “beautiful or pleasant” ranged from approximately 20% (subjects with one brand of implant, rating the oboe) to 85% (subjects with another brand of implant, rating the violin, clarinet and trombone), depending on the instrument and the implant device used by each participant. Only 5% correctly identified the melody and 13.5% correctly identified the instrument name. It is difficult to evaluate these results because it is unclear whether each subject was familiar with each melody and each instrument (some reported little or no musical training).

Fujita and Ito’s study also looked at implanted listeners’ ability to distinguish among five instruments (piano, banjo, violin, harp and trumpet) played on a synthesizer keyboard. Initially subjects could not tell which was being played so they were “trained” to recognize the sounds by receiving feedback after each was played. Testing began when subjects expressed “confidence” in their ability to distinguish among them. Average scores were 56% correct (range 30% to 80%).

A more complete study of instrument identification was recently reported by Gfeller [Gfeller et al., 2002]. The stimuli used in the test included eight instruments: trumpet, trombone, flute, clarinet, saxophone, violin, cello, and piano, each played in three frequency ranges. The recordings consisted of seven-note, connected melodic phrases. For the identification task, subjects were required to select from 16 possible choices. Twenty normal-hearing subjects and 51 implant recipients took this test. The normal-hearing subjects scored an average of 91% correct, while implanted listeners scored 47% on average (a significant difference). The instrument most often correctly identified by implant recipients was the piano. Gfeller noted that most confusions for normal-hearing listeners occurred within instrument family, while implanted listeners tended to show general confusions among all instruments (though no statistics were mentioned to verify this). Another test examining subjective appraisal of “overall pleasantness” of the instruments involved 11 normal-hearing listeners and 48 implant recipients. Mean scores for “likeability” ranged from 60-75 (scale 0-100, range of scores accounts for ratings for different instrument families) for normal-hearing listeners and 40-60 for implant users. However, it is not clear if an exact description of “pleasantness” was given to subjects, and how one’s prior appreciation for an instrument might affect one’s opinion of pleasantness. In the final qualitative test, 24 normal-hearing subjects and 59 implant users were asked to judge the instrument stimuli on bipolar scales: dull-brilliant, compact-scattered, and full-empty. In terms of frequency content, implant recipients judged instruments played in the highest frequency range as sounding significantly more scattered and less brilliant than normal-hearing listeners. Implant listeners also rated the string family as less desirable in tone quality for all three scales.

2.6.2.4 Implant user habits and summary

Gfeller has also assessed implanted listeners’ enjoyment of music and their participation in music-related activities pre- and post-implantation. Using a questionnaire distributed to 65 post-lingually deafened implant recipients [Gfeller et al., 2000], 23% of subjects claimed little satisfaction in listening to music before or after implantation. Forty-three percent claimed that the sound of the music improved with time or that it is better than not being able to hear music at all. Twenty-three percent claimed that music sounds were as pleasant or more pleasant after implantation than before the hearing loss. Also, implant recipients overall seem to enjoy music of a particular genre the same amount as they had enjoyed it prior to implantation. Overall, appreciation of various musical styles decreased from pre- to post-implantation. Subjects were also asked to rate how much they liked the sound quality of 12 instruments they had heard post-implantation. Mean scores for all instruments ranged between approximately 55-70% on average, and there was not one instrument that seemed obviously better than others. Subjects also used a four-point scale to rate their enjoyment level in various musical situations (places of worship, radio, live concert, background music, recorded media, or making music). Overall, the mean ratings indicate that these experiences were not very enjoyable.

To summarize these results, rhythm identification tasks seem the easiest for implant users. Both place and temporal cues can be used to vary pitch perception. Instrument identification seems to be rather difficult for implant users, with a range of 13%-56% correct as compared to 45%-95% for normal-hearing listeners identifying single-tones or musical phrases. Enjoyment of music seems to be compromised by the implant, and qualitative judgements of musical sounds are rather low.

2.7 Summary

Several concepts were introduced in this chapter to provide sufficient background for the current research. The basic physiology of the ear was examined, along with an introduction to hearing loss, cochlear implant technology and simulations. Previous research in the areas of instrument identification by normal-hearing listeners and music perception by cochlear implant recipients was outlined. Normal-hearing listeners with some musical training can achieve fairly high scores in instrument identification, especially when musical phrases are used. Current implants do not seem to provide sufficient information for instrument and melody identification by implanted listeners. We have not found any studies using a simulation of a cochlear implant to examine the impact of cochlear implant processing on musical instrument identification, which prompted the design of the experiment, explained in the next chapter.

CHAPTER 3 METHODOLOGY

The first section of this chapter explains some terminology used throughout the following chapters. Subsequent sections describe the signal processing used to simulate a sound processing strategy used by one implant manufacturer. The real and synthesized audio stimuli are described and analysis is presented which shows the effects of processing on the audio stimuli in time and frequency. Finally, the experiment setup is outlined.

3.1 Terminology

Some terms that will appear frequently in this chapter and beyond are the following:

Unaltered – refers to natural sounds used as test stimuli without having been processed by the implant simulation.

Processed – refers to sounds that have been processed by the implant simulation before being presented to the subjects as test stimuli.

Real instruments – acoustic instruments.

Synthesized instruments – instruments simulated electronically using samplers or synthesizers.

Sampler – a type of synthesizer where instrument sounds are created using samples taken from an acoustic instrument.

Synthesizer – a device that uses mathematical algorithms to create artificial representations of acoustic instruments, aiming to sound perceptually equivalent.

3.2 Simulation

As mentioned in Chapter 2, simulations of cochlear implant signal processing systems have been used before in experiments to assess the implant's ability to convey information for various recognition tasks. Our simulation was modeled closely after Loizou's [Loizou, 1999], with some changes that will be noted as each part of the simulation is outlined. MATLAB was chosen to implement the simulation due to its filtering capabilities, audio functionality and the author's familiarity with the software. The parameters selected for the simulation were based on information available from data sheets and personal communication with Advanced Bionics Corporation™ , manufacturers of the Clarion cochlear implant system.

As the input to the simulation, audio files were imported as .wav files and were resampled to 12971Hz. This step incorporated an anti-aliasing filter as part of the `resample` function in MATLAB. This sampling frequency was chosen as the closest to the 13kHz rate used by the Clarion 1.2 processor, while being an integer fraction of 44.1kHz (which was the sampling rate for most of the input audio files).

The DC component of the input signal was removed. Stereo signals were summed to mono, and the resulting signal was divided into eight channels by seven bandpass filters and one highpass filter (6th-order Butterworth). The bandwidths and center frequencies (the geometric mean of the bandwidth) associated with each channel are listed in Table 1:

Table 1: Bandwidths and center frequencies of bandpass filters

Bandwidth (Hz)	Center Frequency (Hz)
350-494	415.81
494-697	586.79
697-983	827.74
983-1387	1167.7
1387-1958	1648
1958-2762	2325.5
2762-3898	3281.2
3898- (*)	5148.4

(*)The cutoff frequency specified by Advanced Bionics is 6800Hz, which is beyond the Nyquist frequency of our simulation. Thus the cutoff for our simulation was the limit of the anti-aliasing filter from the `resample` function. Nevertheless, the center frequency selected for this channel (displayed in the table) was calculated from the upper cutoff of 6800Hz.

The filters are logarithmically-spaced in frequency to mimic the frequency resolution of the basilar membrane. The Frequency Responses of the filters, shown on linear and logarithmic frequency scales, are displayed in Figure 6.

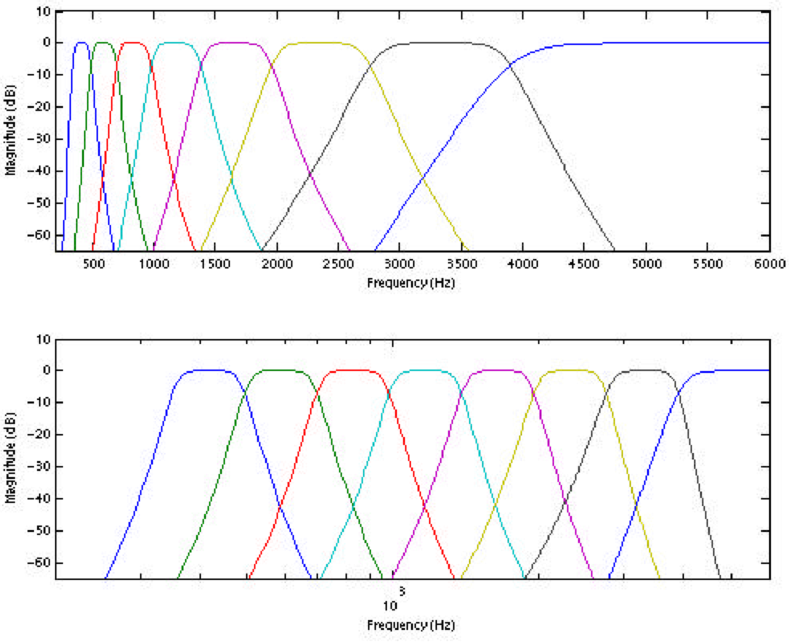


Figure 6: Frequency response of bandpass filters, shown on linear (top) and logarithmic (bottom) frequency scales

Note that the highest frequency filter appears as a high-pass filter, because its specified upper-frequency filter edge is beyond the Nyquist frequency of the resampled signal. In other words, the cutoff of the anti-aliasing will have already removed any frequencies above the specified bandpass edge of 6800Hz.

The next stage in the implant's processing is the extraction of the envelope of the signal from each channel. This is achieved by rectification and lowpass filtering, though other methods exist that may be more effective (i.e. Hilbert Transform [Tierney et al., 1994]). The outputs from each channel are either full- or half-wave rectified, an option that can be configured for each user's preference. Half-wave rectification introduces some low-frequency components due to the fact that components that are harmonics of the original pitch are still present after lowpass filtering; whereas with full-wave rectification, these harmonics get doubled in frequency and are thus beyond the cutoff of the lowpass filter. For both rectification methods, a DC component is introduced, and the harmonics that normally fall above the Nyquist frequency in an analog system are aliased to lower frequencies. The rectified signal is lowpass filtered using a 16th-order moving-average filter, whose frequency response is shown in Figure 7.

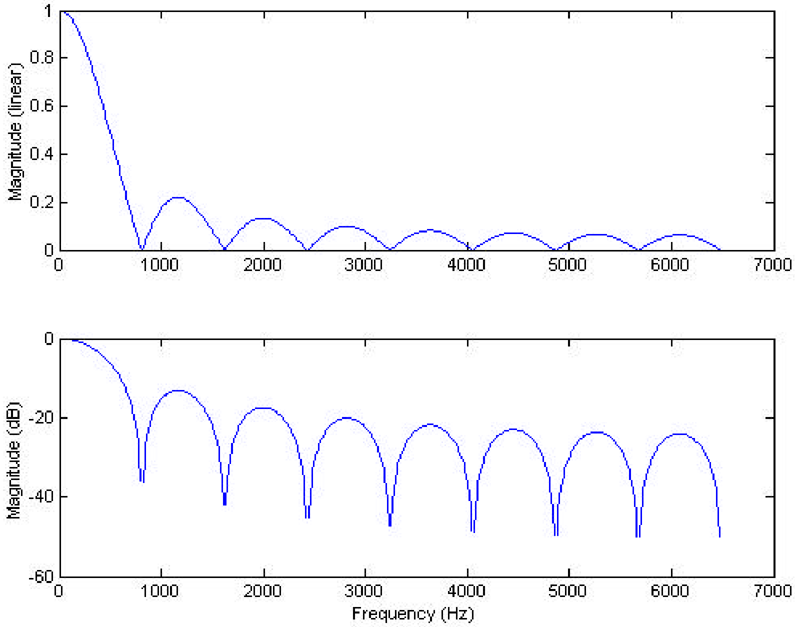


Figure 7: Frequency response of 16-point moving-average filter (linear and dB scale)

Notice that the filter does not have a very sharp cutoff, thus some of the higher frequency components are not eliminated or sufficiently attenuated. The cutoff frequency is meant to be around 400Hz (the -3dB cutoff is 400Hz), but with such high sidelobes, there is some 'leakage' of unwanted high-frequency components. This means that the resulting envelope is not as smooth as it could be, and contains some rapid fluctuations. The Clarion 1.2 device alternatively uses a 6^{th}-order Butterworth filter with a 400Hz cutoff, which results in a smoother envelope. However we were told that the moving-average filter is more commonly used in practice.

In a cochlear implant, the amplitude envelopes of each channel modulate a biphasic pulse train, which has a repetition rate in the range of 800 to 4000 pps. Each modulated pulse train is delivered to a separate electrode, emulating the tonotopic arrangement of the cochlea. The lowest frequency channel delivers its output to the most apical electrode, and the highest frequency channel delivers output to the most basal electrode. In order to enable normal-hearing listeners to hear the amplitude envelopes at frequencies (translating to pitches) similar to those generated by the implant, one must "place" the envelopes on the corresponding locations on the basilar membrane. Thus, the envelopes of the simulation modulate sinusoid carriers, whose frequencies corresponded to the center frequencies of the bandpass filters. For example, an incoming tone at 400Hz would have its envelope modulate a 415.81Hz sine wave, since it falls in the frequency region of the first channel. Since phase is not extracted by current implants, phase information was not incorporated into our simulation. The envelope-modulated sinusoids of each channel are summed together to form the output signal presented to the subjects.

3.3 Sound stimuli

Our interest is the degree to which cochlear implant sound processing limits the identification of musical instruments. In the design of the experiment, one of the questions that arose was whether to use real (acoustic) or synthesized instruments as input stimuli. By using synthesized instruments, factors such as tempo, note duration, pitch and attack could be controlled very precisely. However, using acoustic instruments is more indicative of what occurs naturally, and thus using real instruments has its merits. In order to maximize the information that could be obtained from testing, two experiments were designed, one using acoustic and one using synthesized instruments.

3.3.1 Real (acoustic) instrument signals

The acoustic stimuli for eight instruments were mostly selected from a subset of the database created by Keith Martin [Martin, 1999]. Two of the available four-note phrases were chosen for each of the following six instruments: bassoon, clarinet, cello, flute, oboe and trumpet. The choice of instruments was based on variety in pitch range and similarity in tempo and note duration. Because the existing violin excerpts were different in tempo from the others in the database, new samples were recorded by a soloist in a recording studio. The samples were recorded with a Rode NT2 microphone into ProTools, and were exported as 16-bit monophonic .aiff files. Since piano phrases did not exist, new ones were recorded on a Yamaha upright acoustic piano in the same studio, using a Schoeps microphone (CMC5 preamp and MK4 cardiod head), into ProTools and exported in the same format. The phrases used in experiment training were taken from the database and consisted of various short pieces (roughly 8 seconds long on average) played by each instrument (excluding the piano, for which a new short piece was recorded). All musical signals were normalized to achieve equal loudness in ProTools.

3.3.2 Synthesized instrument signals

In the pilot studies, single-octave major scales were presented to the listeners using the synthesized instruments. However, given the results of the pilot studies, it was determined that these signals were too long in duration, making the experiment too long when testing with several repetitions. Also, the pilot subjects reported that the long phrases gave too much opportunity to "learn" the signals and for one to cue on small nuances of the synthesized instrument notes instead of the characteristic sound of the instrument. For example, with all instruments playing a C major scale, the flute could be recognized distinctly over several notes because of a strange attack that the sampler produced. Thus shorter, three-note phrases were produced.

MIDI note-on commands were generated with the Cakewalk software sequencer. The MIDI notes were then sent to a sampler/synthesizer. Three such samplers/synthesizers were chosen to generate 11 instrument sounds. The choice of synthesizer for each sound was based on how realistic the sound seemed to the author. Table 2 shows the synthesizer that was chosen to generate each instrument sound.

Table 2: Synthesizers used to generate synthesized music sounds

Roland Sound Canvas SC-MKII	E-MU Proteus FX	E-MU Proteus 2 XR
guitar (nylon strings)	cello	bassoon
trumpet	oboe	flute
clarinet	violin	
alto saxophone	trombone	
piano		

The output from the sampler was sent directly to ProTools. The phrases used in the experiment training were single-octave major scales (ascending and descending), either in C or G major, depending on the comfortable range for each instrument. The sounds used for the test were three-note phrases (please see Appendix C for more detail). All sounds were exported as 16-bit, monophonic .aiff files and were normalized to equal volume.

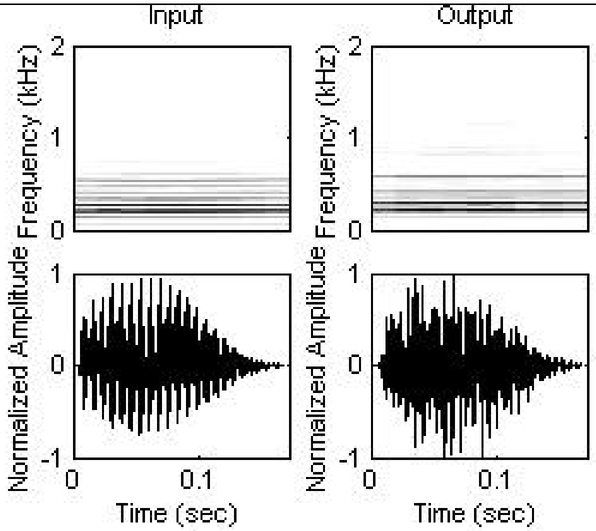
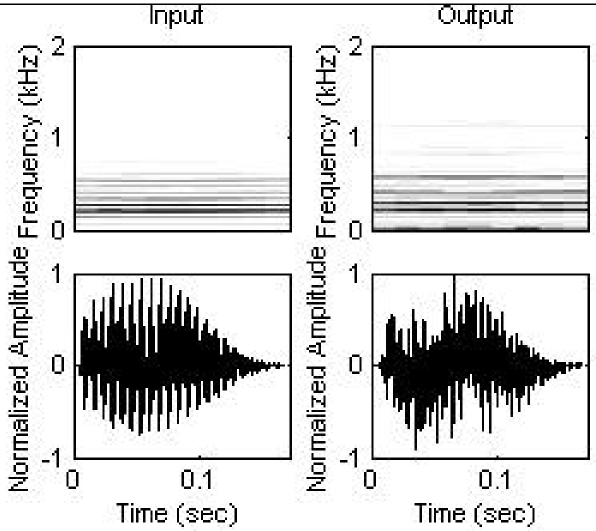
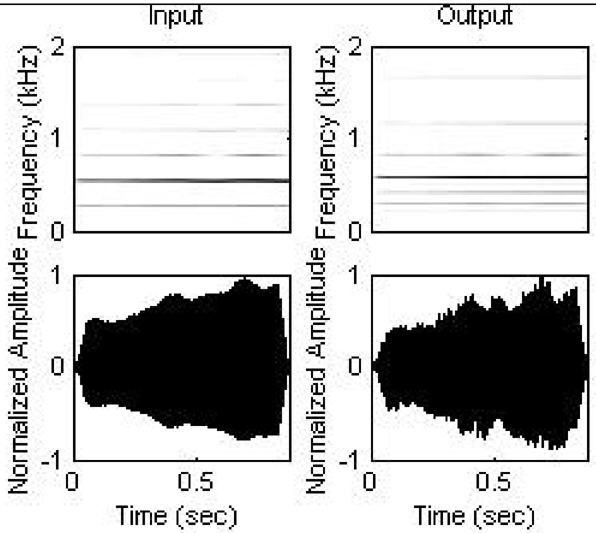
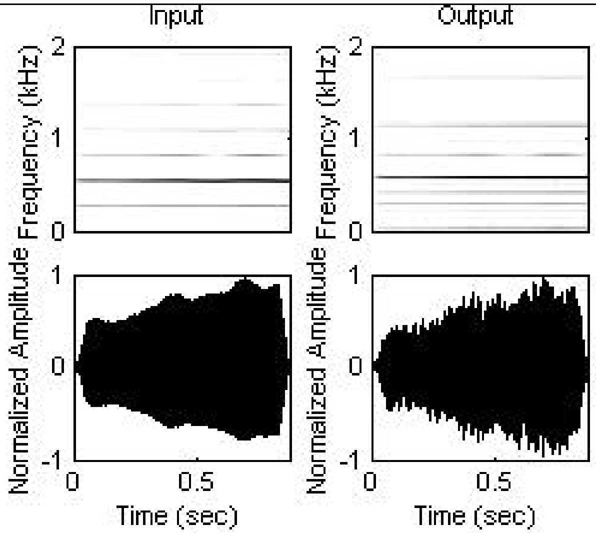
3.4 Example of input and output through simulation

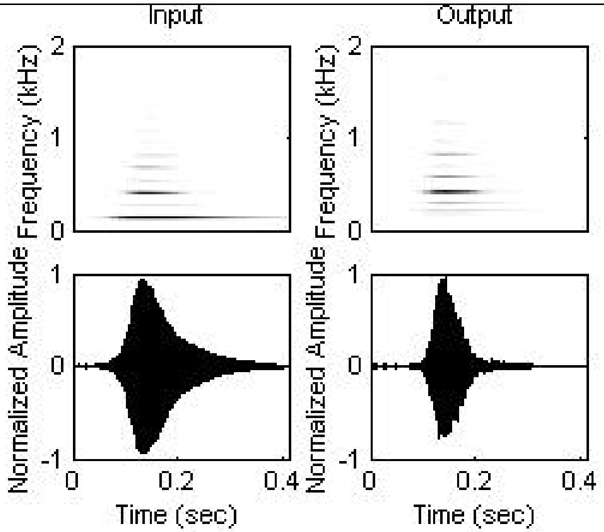
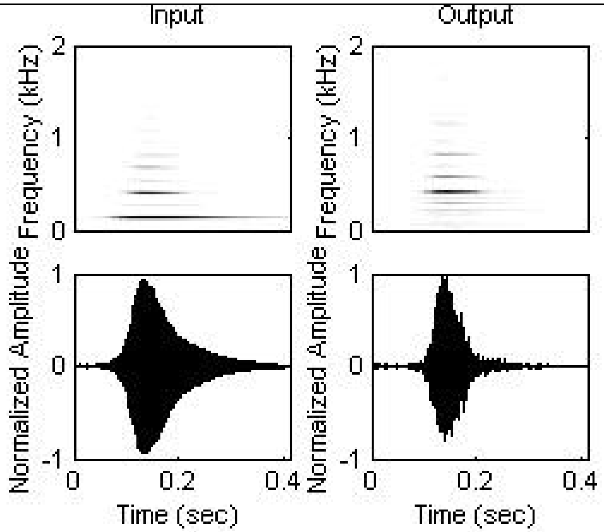
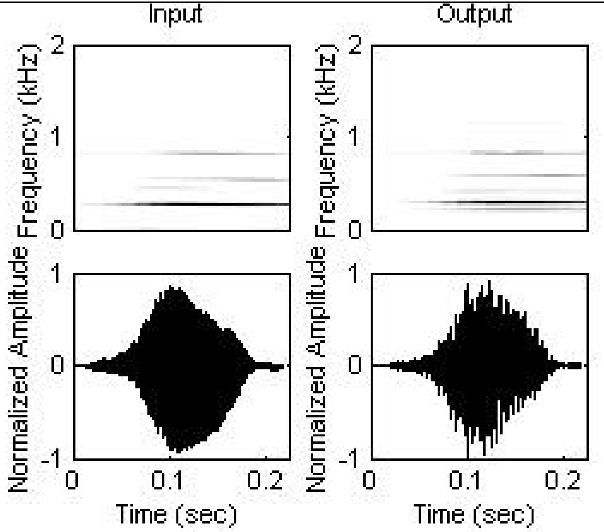
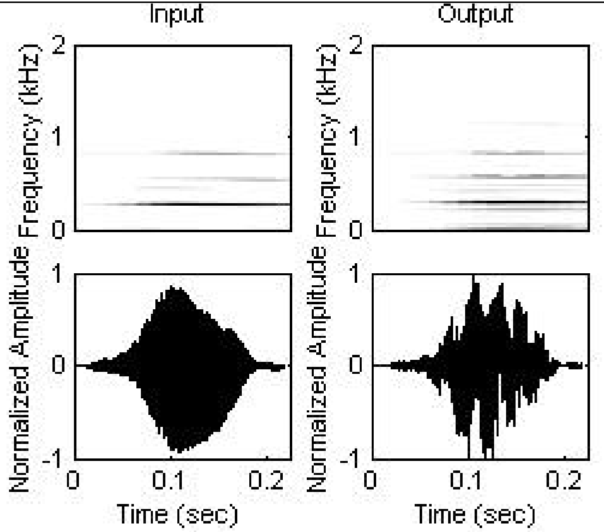
We will now examine the output characteristics of the simulation for all instruments, using single-note inputs played by each instrument.

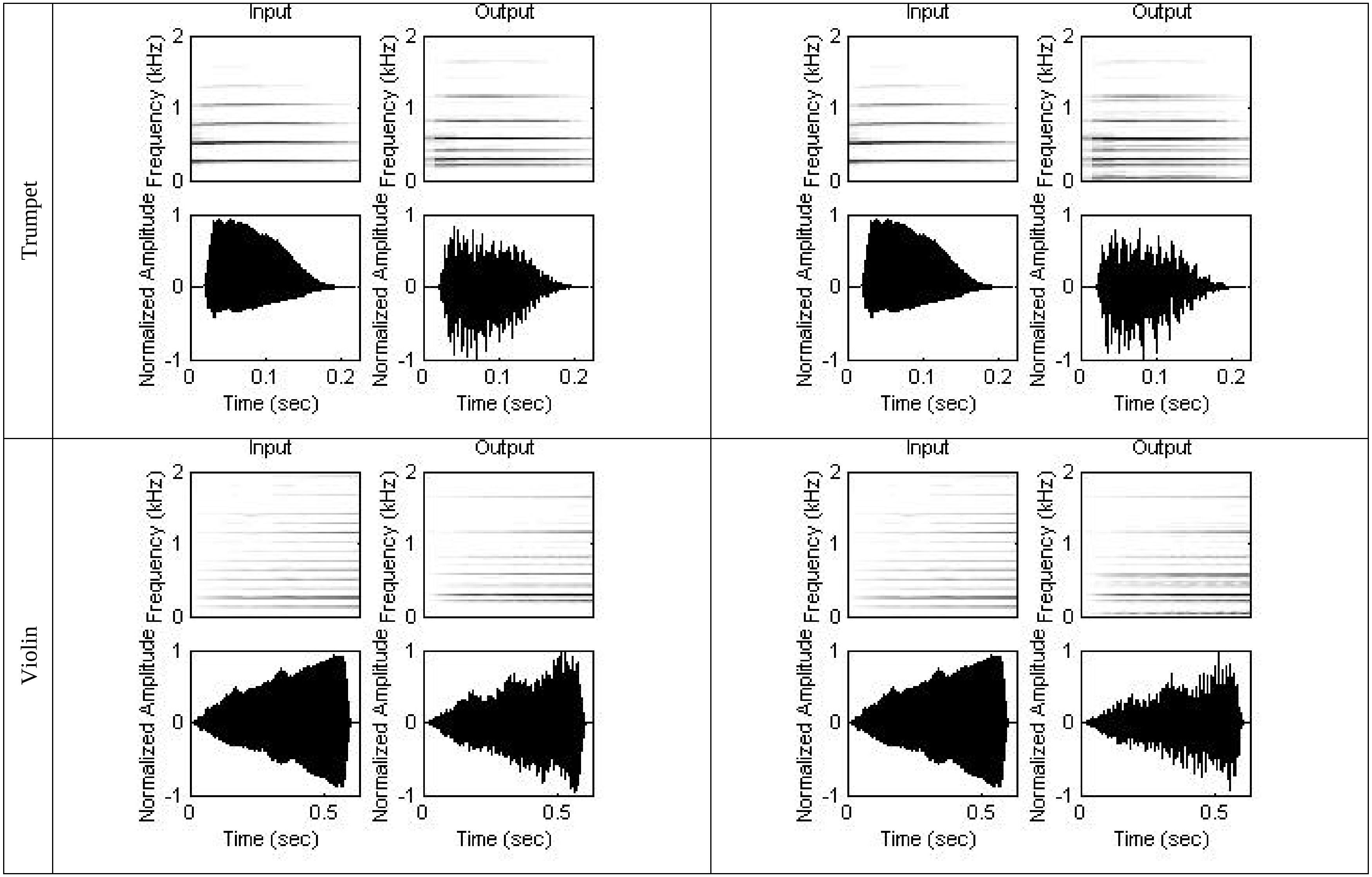
3.4.1 Real instruments

The note chosen for analysis was common to most instruments' recordings, and is approximately C. Recordings from Martin's database were slightly different in pitch than those recorded recently for violin and piano. Flute, trumpet and oboe played C5 (523Hz); piano and clarinet played C4 (261Hz); bassoon and cello played C3 (131Hz).

By plotting the spectrogram of the input and output, we can examine the frequency content of the signals over time. The time waveform allows us to see the changes in the envelope structure after the simulation. Figure 8 shows the spectrogram of the input and output signals, along with the corresponding time waveform representations. Note that both input and output are shown with a sampling rate of 12971Hz, and although higher frequencies are present, the plots are limited to a range of 0-2kHz to show detail. In the figure, the input appears twice to facilitate side-by-side comparison).

	Full-wave rectification	**Half-wave rectification**
Bassoon		
Oboe		

Clarinet		
Flute		


Trumpet
Violin
Input
Output
Input
Output
Frequency (kHz)
2
1
0
Normalized Amplitude
1
0
-1
0
0.1
0.2
Time (sec)
Input
Output
Input
Output
Frequency (kHz)
2
1
0
Normalized Amplitude
1
0
-1
0
0.5
Time (sec)

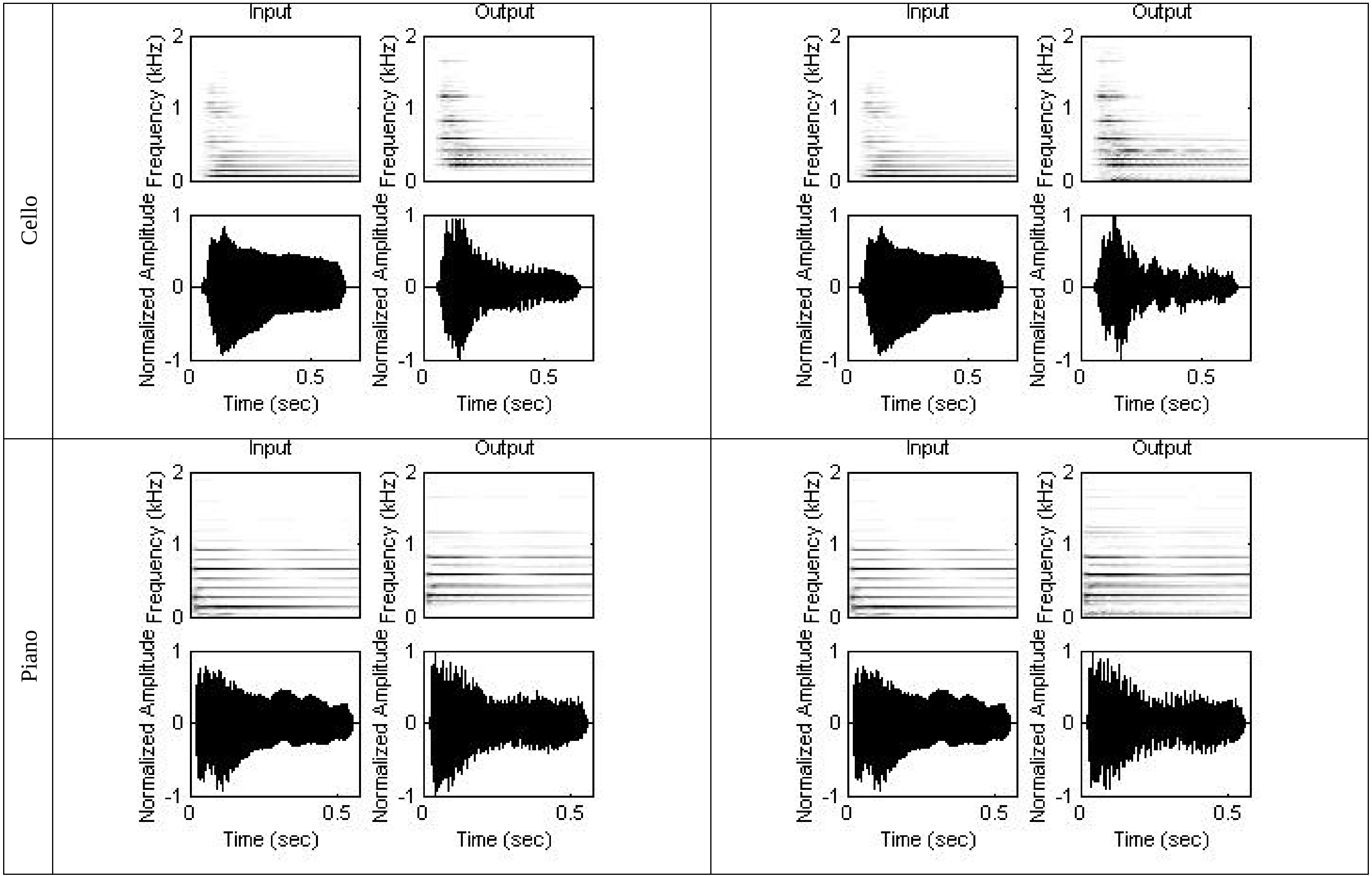


Figure 8: Input and output time waveforms and spectrograms for real instruments

Table 3 presents a summary of the main distortions created by the simulation signal processing with some examples to illustrate these effects for real instruments. For several of these distortions, almost all instruments may be affected, even if only one or two examples are given. Note that for all instruments, the output spectrum is quantized due to the fixed frequency of the sinusoid carrier.

Table 3: Time and frequency domain characteristics of distortions for real instruments, with examples to illustrate each

Time domain characteristics		Frequency domain characteristics	
Description	Example	Description	Example
Distortion of overall envelope shape	The bassoon and trumpet input signals have a smooth, rounded envelope which gets distorted after processing, especially using half-wave rectification.	Alteration in position of energy bands which are no longer harmonics of the fundamental	This effect is very noticeable in the spectrum of the trumpet and violin output signals.
Change in onset (attack) characteristics	The trumpet shows this distortion markedly. The sharp rise of the input during the attack is not conveyed after processing.	Change in weighting of energy bands	The output signal for the cello shows some energy bands with higher frequencies emphasized more strongly than for the input signal.
Change in decay characteristics	The input clarinet signal decays more slowly than the output signal. The input piano signal contains some slow variations during the decay which are not conveyed in the output signal.	Change in range of energy – either low- or high-frequency energy bands are added	Low-frequency energy is added for many instruments, such as the trumpet (especially for half-wave rectification). High-frequency energy bands are added for the piano and cello.
Appearance of rapid noise-like fluctuations appear	Fluctuations can be seen in the output signals for almost all instruments, especially for the oboe, flute, trumpet and piano.	Addition of between-band energy	Even harmonic spacing of the input is not preserved through processing, notably for the oboe (for lower frequencies), trumpet and violin
		Removal of energy bands	The fundamental is the energy band most often removed, occurring for almost all instruments. The output violin signal is missing several energy bands present in the input signal.

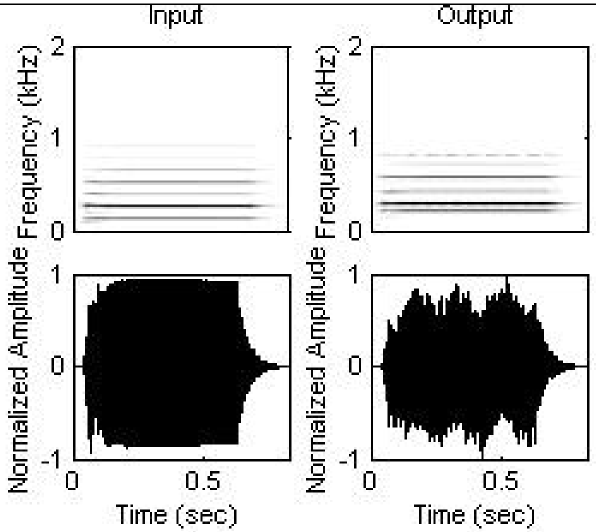
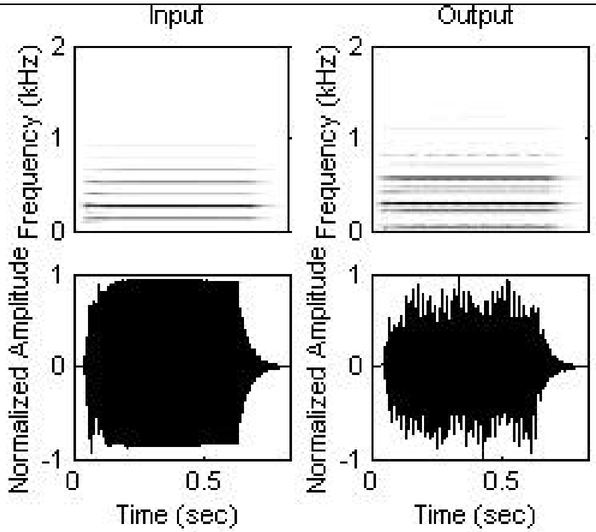
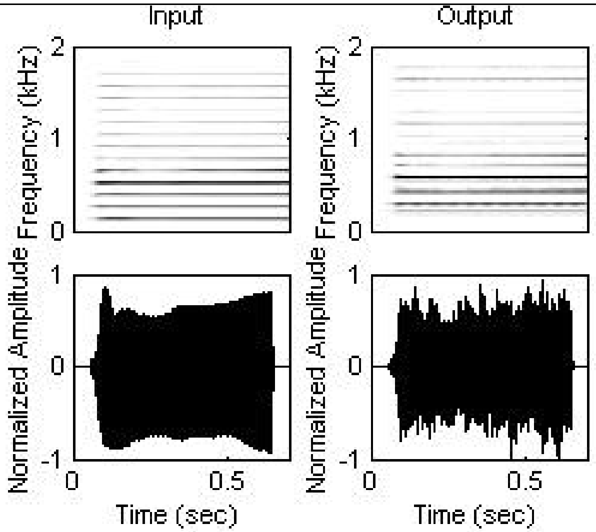
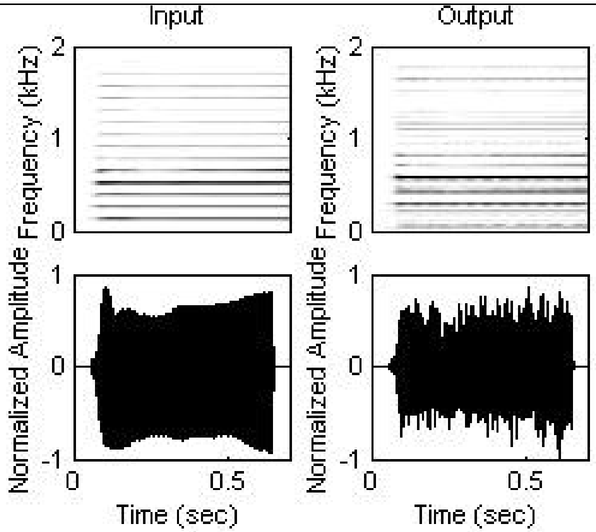
Thus, it seems that all instruments are degraded to some degree in both their time structure (envelope) and spectral content. Distortions can have several sources, as summarized by the table. The fundamental gets removed because of the quantization of the spectrum (i.e., the “new” fundamental becomes the center frequency of the bandpass filter). The addition of high-frequency content is probably a result of the nonlinear rectification and high sidelobes of the lowpass filter. Lower-frequency energy can be attributed to the DC component and aliased high frequencies generated by rectification. The added high-frequeny fluctuations which alter envelope shape can be due to the high cutoff frequency (and high sidelobes) of the lowpass filter. Using visual inspection, it is difficult to visually pinpoint one instrument that is more obviously affected by the processing than another.

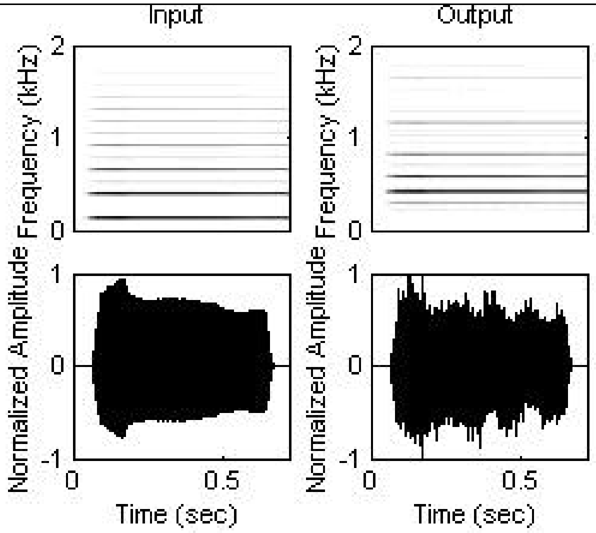
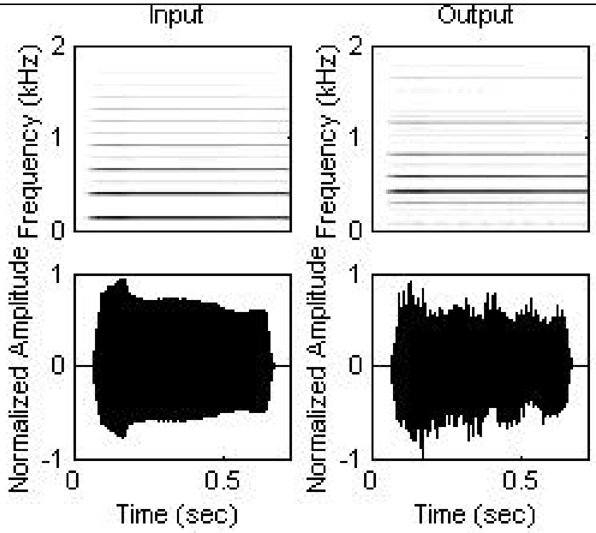
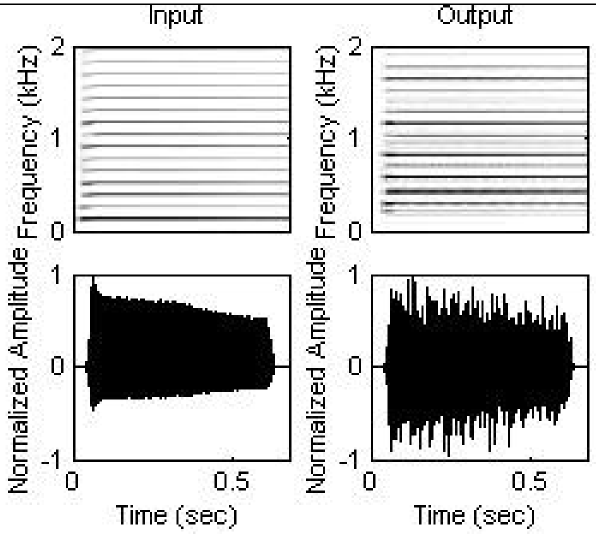
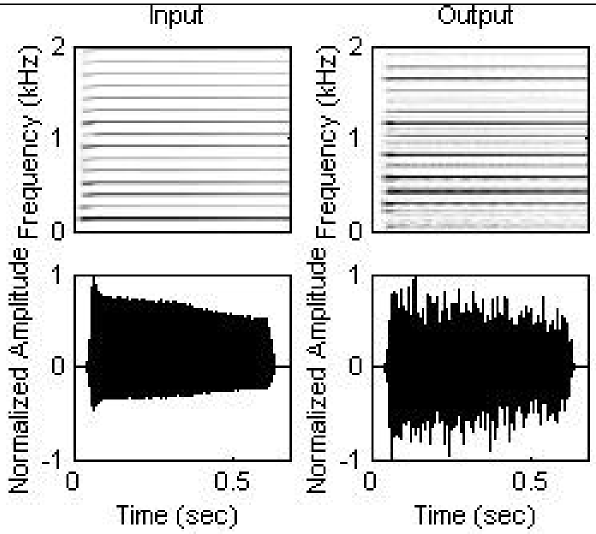
3.4.2 Synthesized instruments

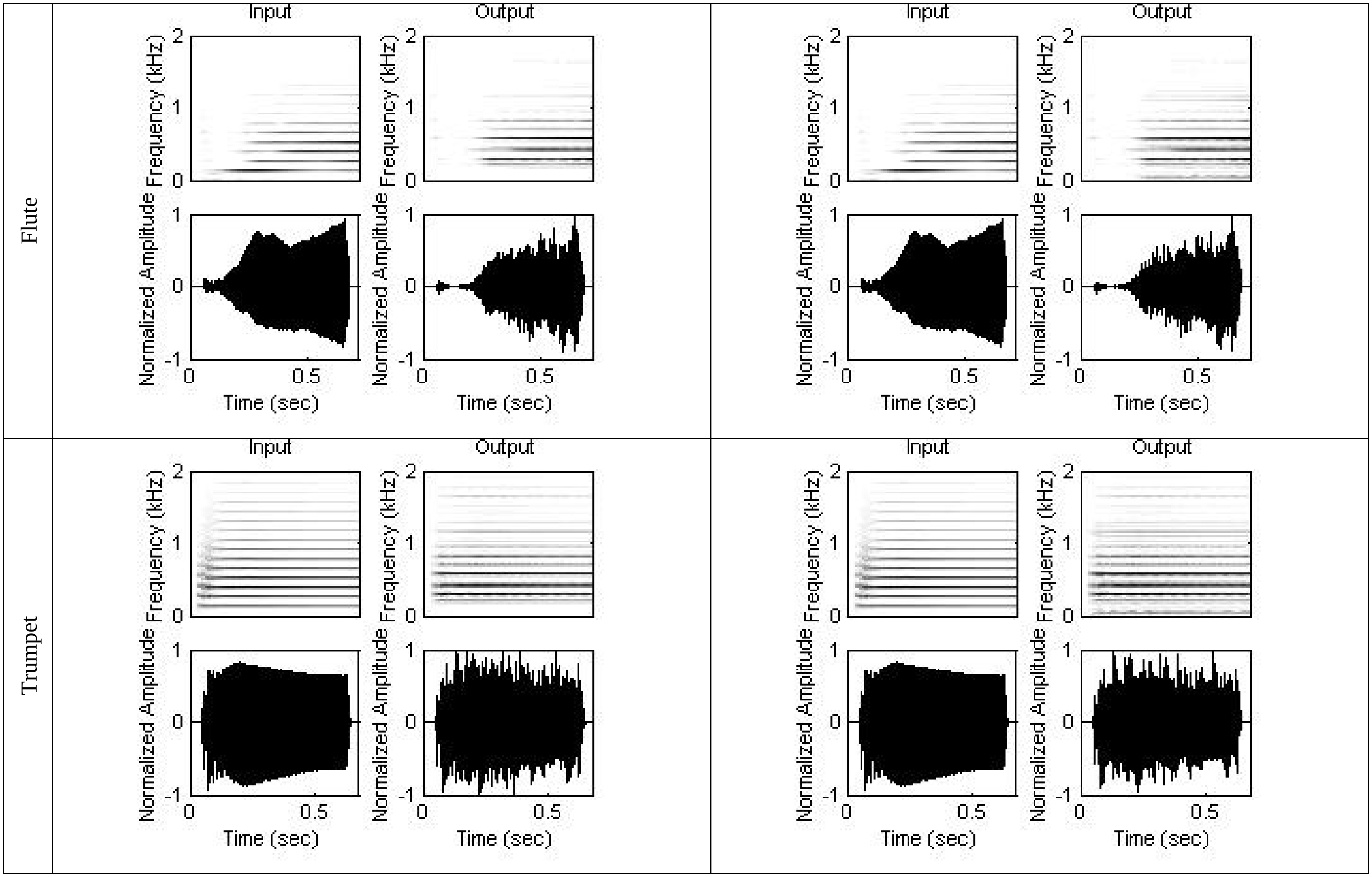
The note C was chosen as the input to the simulation for visual examination of the distortions produced for all synthesized instruments. C4 (261Hz) was played by saxophone, bassoon, oboe, clarinet, flute, trumpet, piano, cello and violin. C3 (131Hz) was played by guitar, and C2 (65Hz) was played by trombone. Figure 9 shows the spectrograms and time waveforms for these instruments, before and after processing by the simulation. Table 4 presents a summary of the main distortions created by the simulation signal processing with some examples to illustrate these effects for synthesized instruments. For several of these distortions, almost all instruments may be affected, even if only one or two examples are given.

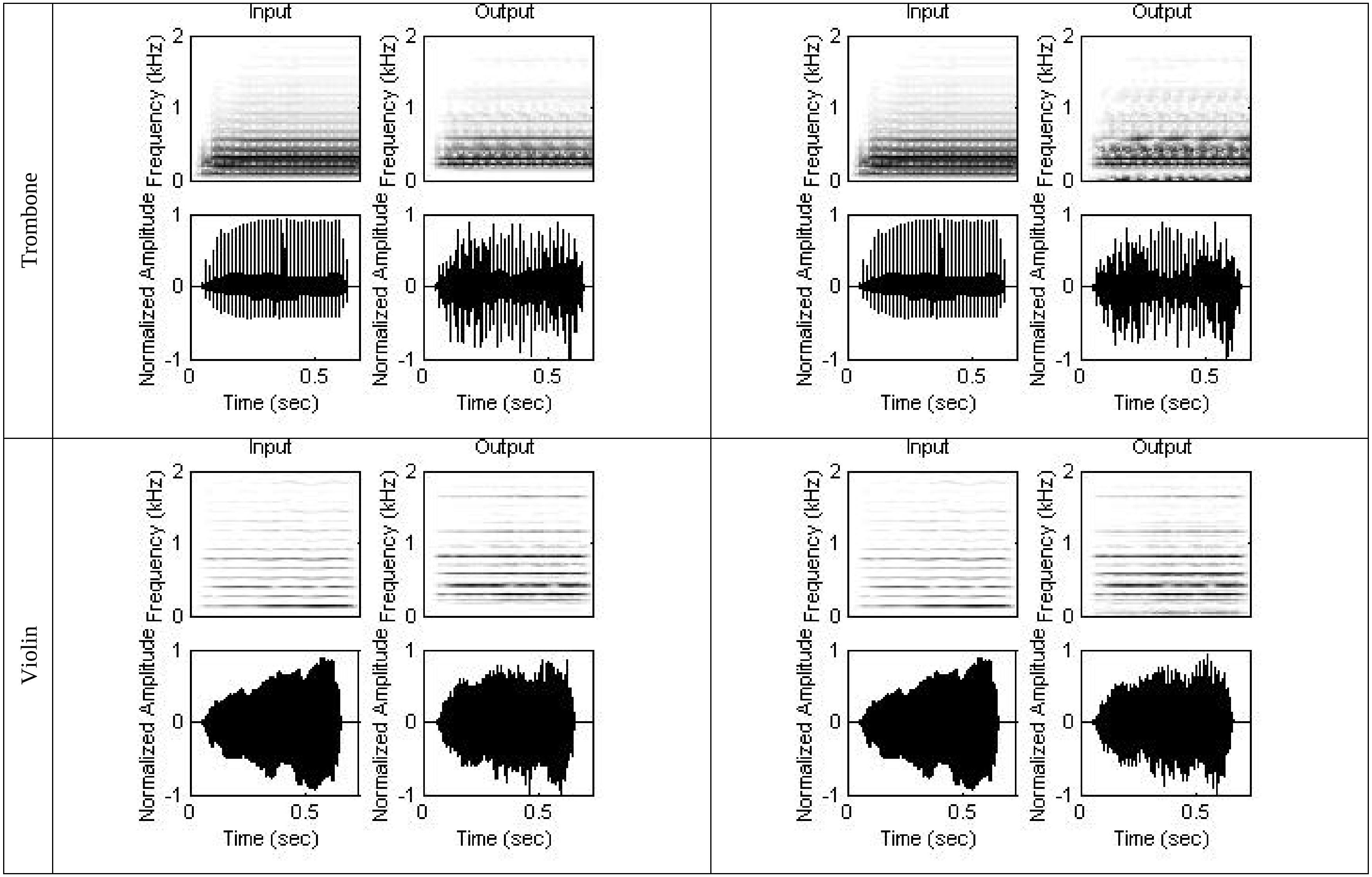
Table 4: Time and frequency domain characteristics of distortions for synthesized instruments, with examples to illustrate each

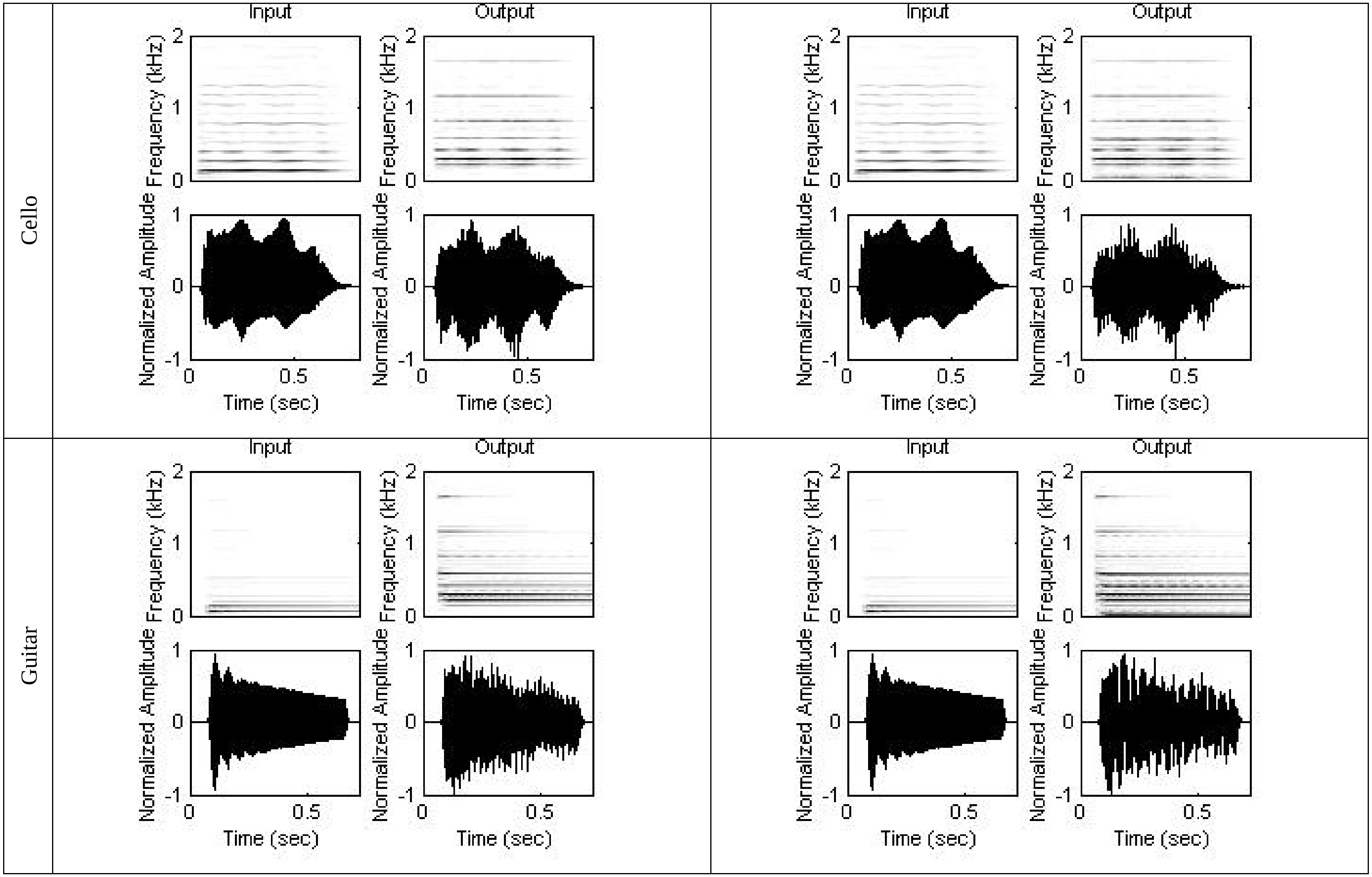
Time domain characteristics		Frequency domain characteristics	
Description	Example	Description	Example
Distortion of overall envelope shape	Several instruments exhibit this distortion, such as the bassoon and trombone.	Alteration in position of energy bands which are no longer harmonics of the fundamental	The input oboe signal contains very uniformly spaced harmonics while the output shows energy bands that are not multiples of a fundamental (especially for half-wave rectification).
Change in onset (attack) characteristics	The input oboe, saxophone and piano signals have a sharp rise followed by a quick decay, which is not conveyed in the output signal.	Change in weighting of energy bands	The output saxophone signal exhibits high-frequency energy bands with more emphasis than in the input signal.
Change in decay characteristics	The smooth decay of the bassoon is distorted in the output. The decay of the guitar does not descend as gradually or uniformly as in the input.	Chance in range of energy – either low- or high-frequency energy bands are added	The input guitar signal contains mostly lower frequencies, but the output contains new energy at higher frequencies.
Appearance of rapid noise-like fluctuations appear	Most instruments exhibit this property quite prominently.	Addition of between-band energy	Almost all instruments exhibit this distortion. The saxophone, for example, shows non-uniformly-spaced energy bands compared with the evenly-spaced harmonics of the input.
		Removal of energy bands	In most cases, the fundamental is removed (the clarinet shows this for full-wave rectification).

	Full-wave rectification	**Half-wave rectification**
Bassoon		
Oboe		

Clarinet		
Alto Saxophone		


Flute
Trumpet
Input
Output
Frequency (kHz)
Normalized Amplitude
Time (sec)
2
1
0
-1
0.5


Trombone
Violin
Input
Output
Frequency (kHz)
Normalized Amplitude
Time (sec)
2
1
0
-1
0.5


Cello
Guitar
Input
Output
Frequency (kHz)
Normalized Amplitude
Time (sec)
2
1
0
-1
0.5

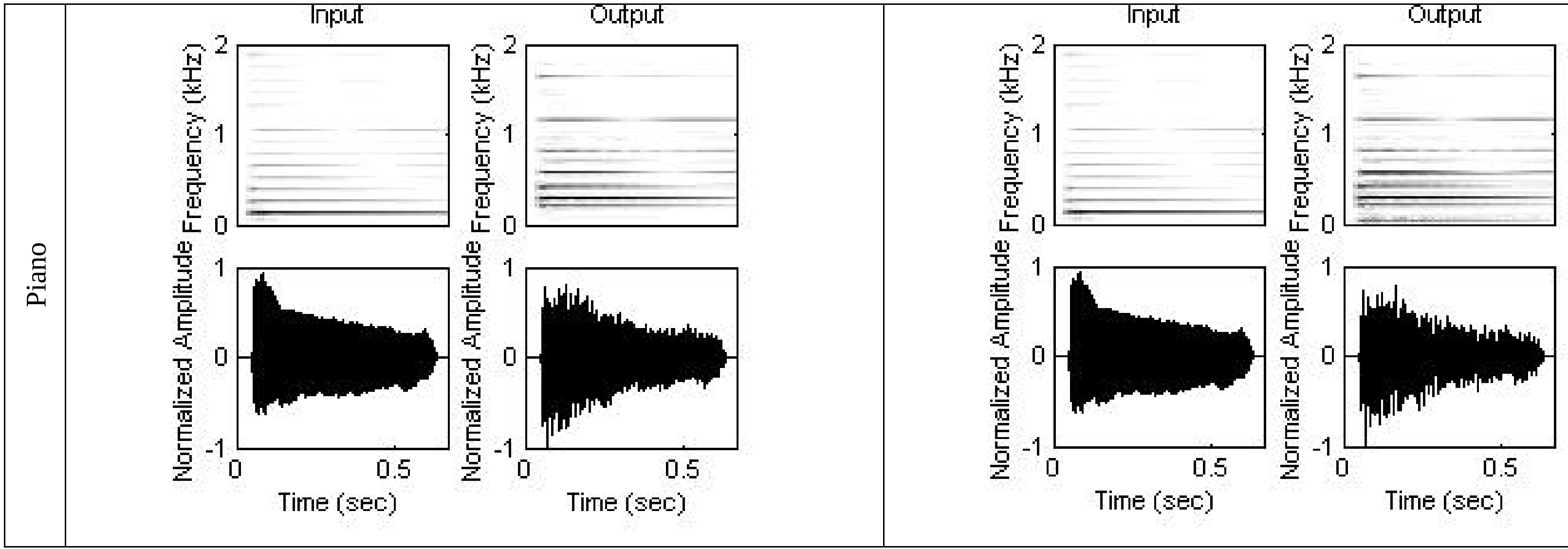


Figure 9: Input and output time waveforms and spectrograms for synthesized instruments

As was the case with real instruments, the distortions of the simulation can be seen in the time and frequency signals of the output of the simulation for synthesized instruments. It is difficult to visually assess the degree to which one instrument signal is more distorted than another. The listening experiments described next are designed to determine to what extent these distortions affect the ability of normal-hearing subjects to identify these instruments.

3.5 Experiment design

Several iterations of experiment design were considered and tested. Some of the issues encountered involved how to train the subjects and how to present the stimuli. It was clear that it was necessary to test both the processed and unaltered sounds: the latter informing us of a subject's 'baseline' ability to identify the instruments, and the former providing information about the degree to which the distortions created by the simulation affect this ability. One issue was the possibility for subjects to hear some nuance in the unprocessed recording (e.g., a breath, 'click' from fingering the instrument, or a particular bowing release) that they would identify in the processed phrase. While this applies mostly to real instruments, similar cueing effects could occur with synthesized instruments. In order to reduce the likelihood of such effects, the sounds used in the training were different from those heard in the test segments. Also, only unaltered sounds were played in the training, and processed sounds were heard first in the testing procedure.

There were three segments to the overall testing procedure: (1) training, (2) identification of processed sounds (called "Part I") and (3) identification of unaltered sounds (called "Part II"). The three segments formed one complete testing procedure. Every subject performed the three-segment test twice: once for real instruments, and once for synthesized instruments. Each of the two testing procedures lasted approximately 30 minutes, thus the total testing time was close to one hour. The order of presentation of the tests for real and synthesized instruments were counterbalanced; that is, half of the subjects tested on real instruments first, while the other half began with synthesized instruments. This ensured that any influence of learning was counterbalanced across the real and synthesized conditions. Each of the sections of the testing procedure is now outlined in detail.

3.5.1 Training protocol

The purpose of the training was to familiarize the subjects with the testing software, and to expose them to each of the unaltered instrument sounds. The instrument stimuli were presented one at a time. The subject could replay the sound as many times as s/he liked, and was told which instrument was playing. There were the same number of instrument choices as there were instrument stimuli. The subject was required to select an instrument from a list of possibilities and submit this choice in order to hear the next stimulus. Even though they were told which instrument was playing, the selection process was required to familiarize the subject with the software. The experimenter was present during this time to demonstrate the software and answer any questions.

3.5.2 Part I (processed sounds) protocol

3.5.2.1 Real instruments

Two 4-note phrases were selected for each of the eight instruments. Each phrase was played three times over the course of the experiment. Each of the phrases was processed twice by the simulation prior to the experiment; once using full-wave and once using half-wave rectification. Thus, for this part of the experiment there was a total of 96 stimuli (2 phrases x 3 repetitions x 2 processing conditions [full- and half-wave] x 8 instruments). The stimuli were pseudo-randomized such that any one instrument would not be heard twice in succession. Twenty randomized “playlists” were generated before subjects were enlisted for the experiment, and when a subject took the test, one of the 20 lists was randomly selected. Subjects heard the stimulus once and were instructed to select one instrument from the list of all eight instruments. There was no time limit for the selection process and feedback was not given.

3.5.2.2 Synthesized instruments

Three 3-note phrases were generated for each of the eleven instruments (please refer to Appendix C for more detail). Each phrase was played twice over the course of the experiment for both the full- and half-wave rectification conditions. For this experiment, the total number of stimuli was 132 (3 phrases x 2 repetitions x 2 processing conditions [full- and half-wave] x 11 instruments). The randomization of stimulus presentation and the response technique were the same as described for the real instrument condition.

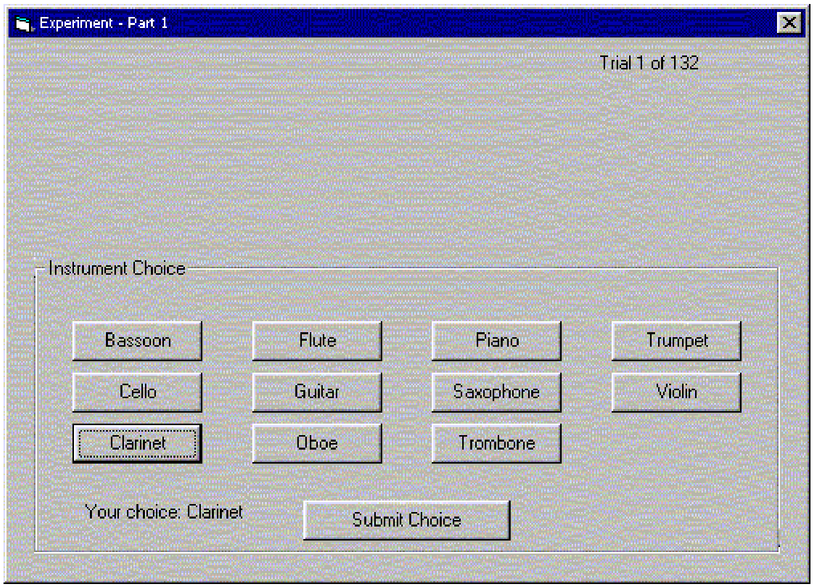
3.5.3 Part II (unaltered sounds) protocol

The experimental setup was the same as described for Part I, but with half the number of input signals since there were no longer two processed conditions (stimuli were in their natural form). The stimuli were presented at 48kHz for real instruments (44.1kHz for synthesized instruments) and were randomized in the same manner as previously described.

After completion of the experiments for real and synthesized sounds, the subjects completed a questionnaire and musical background survey (see Appendix B).

3.6 Experiment software

The testing software was written in Visual Basic. Each experiment (real and synthesized) was its own executable program. The testing screen, shown for Part I of the experiment using synthesized instruments, is displayed in Figure 10.



Trial 1 of 132

Your choice: Clarinet

Figure 10: Screen shot of testing software (Part I, synthesized instruments)

The buttons are the instrument choices and only become active when the sound finishes playing. The subject can press any instrument and the selection will be displayed next to the “Your choice” label. The “Submit Choice” button is pressed when the subject is sure of the answer s/he selected, which then queues the next stimulus which is played automatically. The current trial number is displayed in the upper righthand corner. The instruments that were played and selected are written to a file, and the filenames of the stimuli used in each part of the experiment are written to a separate file. The files with the selections are later imported into MATLAB for analysis.

3.7 Subjects

Recruitment of subjects was done by e-mail and by posted flyers on the MIT campus. Potential subjects were told that they should be able to identify all of the 11 instruments in the experiment by hearing a short audio clip. Advertisements for the experiment also mentioned that the subjects would be compensated $10 upon completion of the two experiments (real and synthesized instruments). A total of 27 subjects took the test. Subjects were instructed to place themselves into one of four age categories, resulting in the following demographics: 15 subjects were 17-30 years old, ninebetween 31-50 and three aged 51 or older. All except two reported no known hearing problems (one of the two claims difficulty in hearing a speaker in a multi-talker situation, and the other claims some hearing loss from prolonged exposure to playing percussion). The range of musical background was rather varied. Subjects indicated if and for how long they have played and/or received instruction on an instrument, as well as number of years performing with an ensemble and any courses (music appreciation, theory, ear-training, etc.) taken. Six of the subjects were music majors in college. The subjects classified themselves into one of five categories (taken from Gfeller [Gfeller, 2000]):

1. No formal training, little knowledge about music, and little experience in listening to music
2. No formal training or knowledge about music, but informal listening experience
3. Self-taught musician who participates in musical activities
4. Some musical training, basic knowledge of musical terms, and participation in music classes or ensembles in elementary or high school
5. Several years of musical training, knowledge about music, and involvement in music groups

The mean value for the self-assessment was 4.31, with a standard deviation of 0.93. Subjects signed a consent form and answered questions relating to the experiment and musical background (please see Appendix B). The approval for use of human subjects was covered by the MIT’s Committee on the Use of Humans as Experimental Subjects (COUHES) application 2862.

3.8 Testing setup

Each subject sat in a quiet, sound-isolated recording studio. The testing software was presented on a PIII-450MHz laptop. The subjects were first asked to indicate if the listening volume was comfortable, using the synthesized piano major scale as a sound file. Subjects listened over AKG K240 headphones. They were briefed on the experiment and were asked to sign a consent form. They were told that the sounds they were going to hear were not necessarily meant to represent exactly what someone with an implant actually hears, but are used to assess a normal-hearing person's ability to distinguish among different musical instruments, after the original signals have been passed through the simulated processor. Subjects took the tests for real and synthesized instruments separately with a short pause in between the two (the order of the experiments was counterbalanced for each subject). Total testing time was approximately one hour per subject.

3.9 Summary

This chapter began with a description of the MATLAB procedure used to simulate the sound processing strategies used by the cochlear implant. Audio stimuli were generated by real and synthesized musical instruments, and these signals were processed by the simulation to obtain the output sounds. A representation of each instrument before and after the processing was shown in both the time and frequency domains to demonstrate the types of distortions that occur. The experimental setup was outlined as well as the demographics of the 27 subjects who participated. The next chapter will present and discuss the results from the listening experiments.

CHAPTER 4 RESULTS AND ANALYSIS

This chapter presents the results obtained from testing the normal-hearing subjects on the instrument identification tasks. As explained in the previous chapter, subjects were trained on one set of unaltered instrument sounds and were asked to identify a different set of processed and unaltered sounds, for both real and synthesized instruments.

Of the 27 subjects who completed the experiments, two subjects' data were removed since only results for one of the two experiments was recorded due to a problem with the software. For all results shown in this chapter, the instruments are identified by numbers as shown in the following table.

Table 5: Legend of instrument number and name

Instrument #	Name
1	Bassoon
2	Oboe
3	Clarinet
4	Alto saxophone
5	Flute
6	Trumpet
7	Trombone
8	Violin
9	Cello
10	Guitar
11	Piano

The results of analyses of various groups of subjects are presented in this chapter in order to directly address our main question: to what extent the processing of the implant affects identification of musical instruments by musically-trained listeners. First, all subjects' data will be analyzed. The second group for analysis consists of those who scored highest for unaltered real instruments and the third group consists of those who scored highest for unaltered synthesized instruments. The effect of order of experiment will be examined (half of the subjects performed identification for real instruments first, while the other half identified synthesized instruments first). We present the results for the real instruments followed by results for the synthesized instruments.

4.1 All subjects

The following analysis applies to all 25 subjects with complete results. Some subjects seem better at identifying the instruments than others. The mean identification scores are plotted against the information that the subjects provided in self-assessing their musical abilities (scores range from 1-5, higher numbers indicating more formal musical training). Figure 11 shows this information for both real and synthesized instruments.

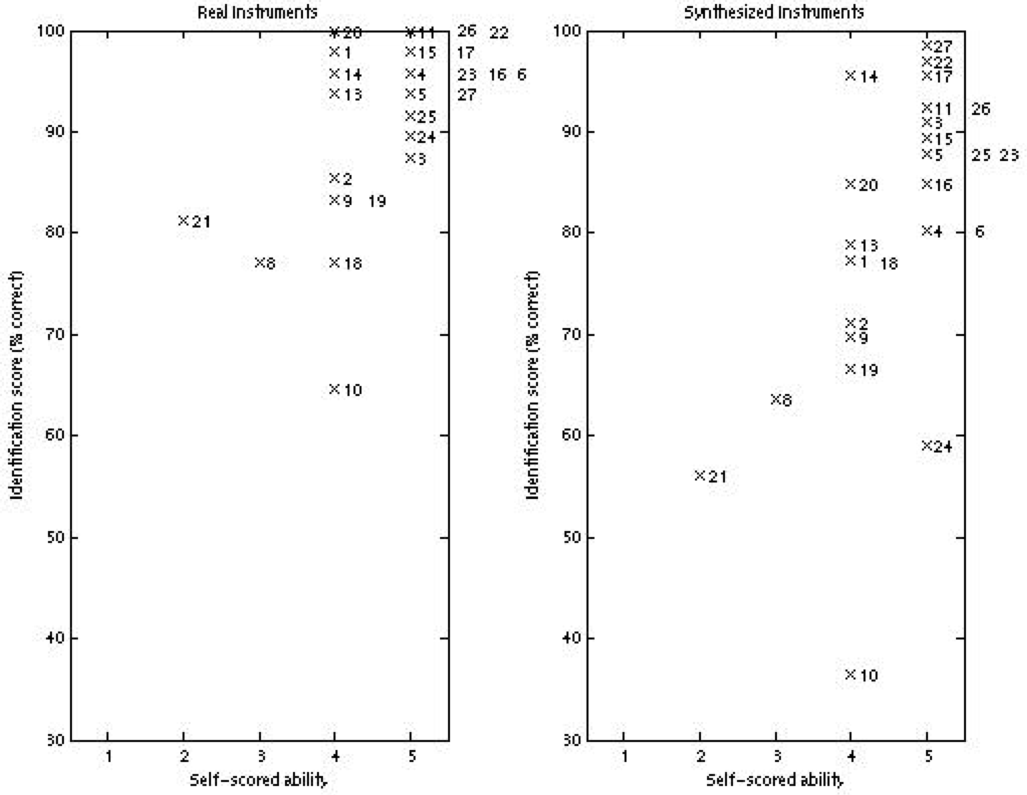


Figure 11: Mean instrument identification scores of unaltered sounds for all subjects (indicated by subject number) as a function of their self-assessed musical ability. Subjects who scored the same for identification and musical ability are represented by one symbol followed by their subject numbers (offset to the right).

Note that some subjects scored the same for identification and self-assessed ability. Rather than showing these subject numbers overlapped, they are represented by a single symbol with their subject numbers offset to the right. For example, for real instruments, subjects 9 and 19 both rated themselves as having ability ‘4’, and both obtained a mean score of 83.3% correct. For real instruments, the following groups of subjects (shown in round parentheses) obtained identical mean scores for a given ability score [shown in square brackets]: [(9,19) for ability ‘4’], [(5,27), (4,23,16,6), (15,17), (11,26,22) for ability ‘5’]. For synthesized instruments, the following groupings occurred: [(1,18) for ability ‘4’], [(5,25,23), (11,26) for ability ‘5’]. Note that there are a number of occurrences of low scores for subjects who rated their ability “high”. For example, subject 10 self-assessed as having ability ‘4’ but scored lower than all subjects for both real and synthesized instruments; subject 24 with ability ‘5’ scored poorly for synthesized instruments. Also note that subject 21, who indicated ability ‘2’, scored better than three other subjects with supposedly higher ability. The correlation for the two variables is r = 0.57 (significant for p = 0.01) indicating a moderate degree of relation for this range of data. These results show that one can only partially rely on a subject's self-assessed mean to predict his/her ability to identify instruments.

Overall, the distribution of scores was lower for the synthesized instruments, even for musically-able subjects. The variables were correlated at r = 0.61 (significant for p = 0.01), which indicates a higher and slightly stronger correlation than for real instruments. These mean scores included scores for all instruments; the next sections describe the results specific to each of the instruments played, for all subjects.

4.1.1 All subjects - real instruments

Figure 12 shows identification scores for each instrument and each subject. For each instrument, a mean score is shown for unaltered sounds, processed sounds using full-wave rectification, and processed sounds using half-wave rectification. The numbering of the instruments appears non-sequential since there were no real instrument stimuli for the saxophone (4), trombone (7) or guitar (10).

There is considerable variability across instruments in the subjects' ability to identify the unaltered and processed instrument sounds. For unaltered sounds, the variability in musical background may be a factor, as was seen in Figure 11 (recall that self-assessed ability and identification scores were only moderately correlated). It could also be a result of differences in stimuli for each instrument, or whether a particular subject was more familiar with one instrument over the others.

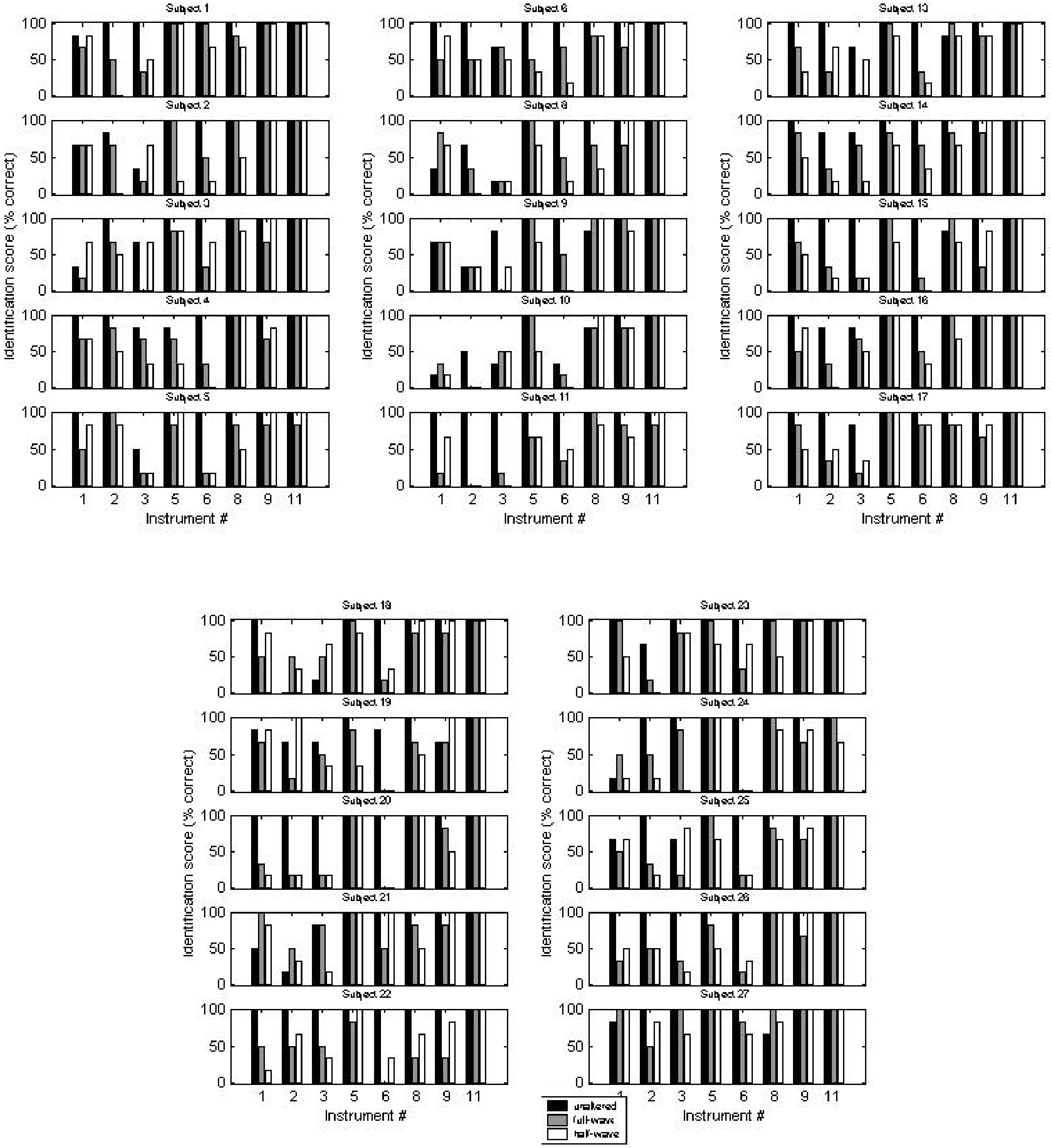


Figure 12: Identification scores for each subject using real instrument sounds. Mean scores for each instrument are shown for unaltered sounds, processed sounds using full-wave rectification, and processed sounds using half-wave rectification.

The average scores for all subjects for both unaltered and processed stimuli are shown in Figure 13, along with the error bars corresponding to the standard deviation. The processed scores include scores for both full-wave and half-wave rectification methods.

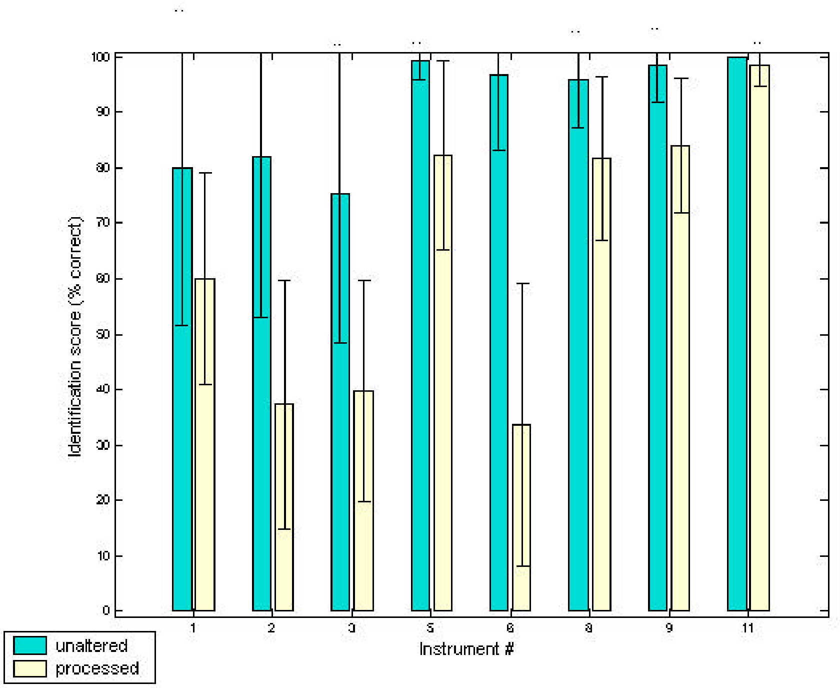


Figure 13: Mean scores across all subjects for each instrument testing unaltered and processed real instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

The question we wish to answer is how the processing affects the subjects' ability to identify each input instrument sound and to what extent the processing affects some instruments more than others. After plotting the histograms of scores for each instrument across all subjects, it was apparent that they were not consistent with a normal distribution. Thus, the non-parametric, Wilcoxon rank sum test was used to compare the score distributions obtained for the unaltered and processed cases for each instrument. Processing resulted in significantly lower scores (p=0.01) compared with unaltered sounds, for all instruments except the piano (Instrument #11, written as (11) from this point onward).

When comparing results for the full- and half-wave rectification conditions, the difference in scores was significant (p = 0.01) for flute (5), violin (8) and cello (9). For flute and violin, mean scores for full-wave were higher than half-wave. The cello was recognized more often by all listeners when half-wave rectification was used. This might be due to the lower frequency "rumbling" sound that the half-wave rectification introduces, which may be more consistent with the lower-frequency sounds of the cello. Figure 14 shows average scores for full- and half-wave rectification.

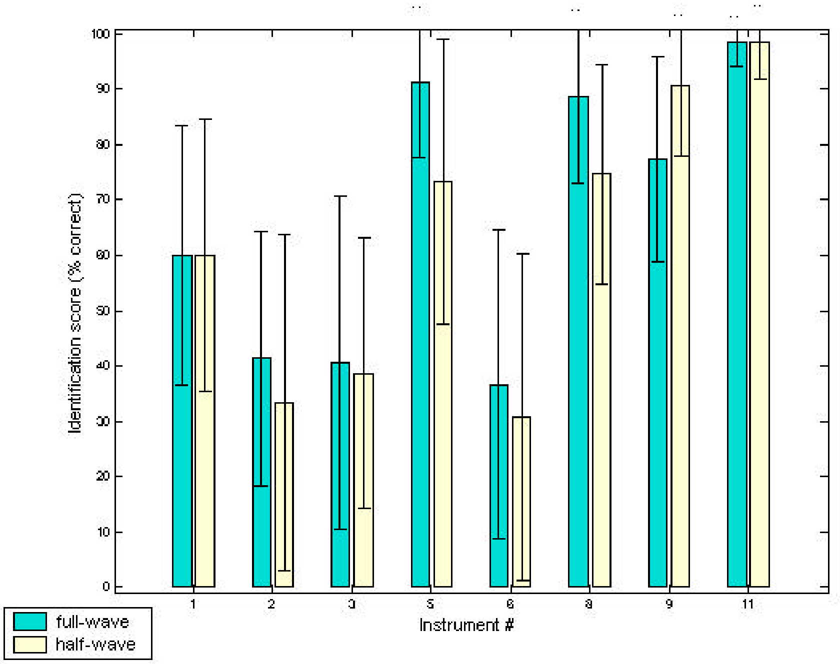


Figure 14: Mean identification scores across all subjects for each instrument on processed real instrument sounds, grouped by full- and half-wave rectification

We are interested in the extent to which each instrument is affected by processing and would like to know if some instruments are more robust to processing than others. In order to calculate this, the difference between unaltered and processed scores for each subject was determined, for each instrument. Note that in some rare instances, a subject might score higher on the *processed* sounds, possibly due to chance, thus a negative difference may be obtained. We then plot the distribution of the differences in scores across all subjects for each instrument. The mean of these distributions is then calculated, and the ranking is ordered according to the mean. Thus, the lower the mean difference, the less an instrument was affected by the processing. The following figure shows the distribution of difference scores across all subjects for each of the real instruments.

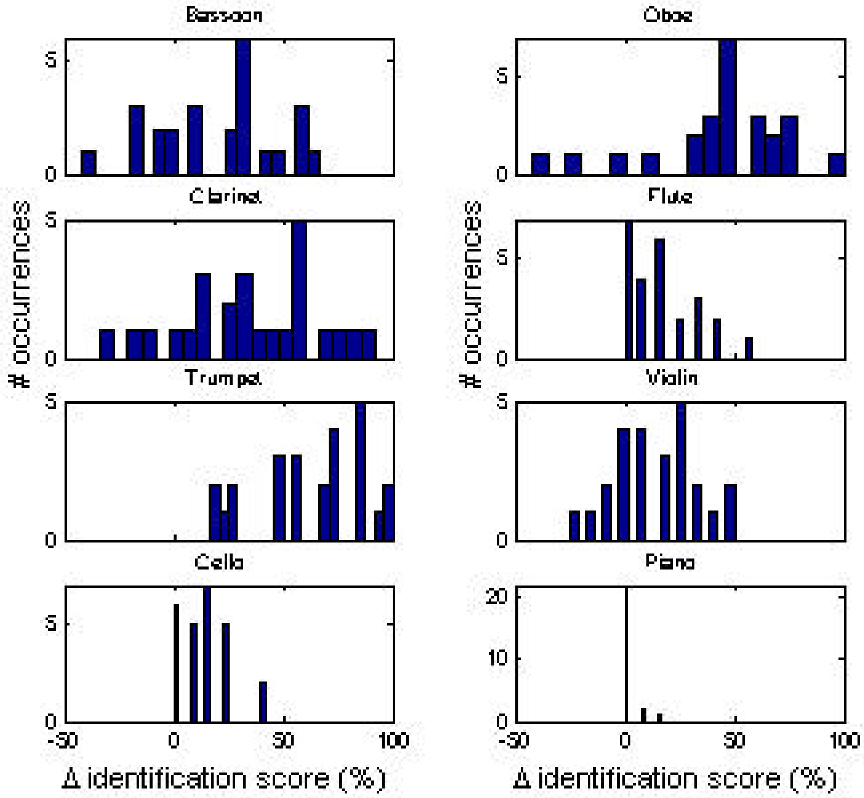


Figure 15: Frequency of occurrence of differences in identification scores between unaltered and processed sounds, across all subjects, for each real instrument.

The mean difference and standard deviation in identification scores for real instruments are shown in Figure 16, which illustrates the increasing order of difficulty in identification (i.e., increased distortion) of each instrument. The woodwinds grouped together, and are more distorted than strings or piano. The trumpet is most affected by processing.

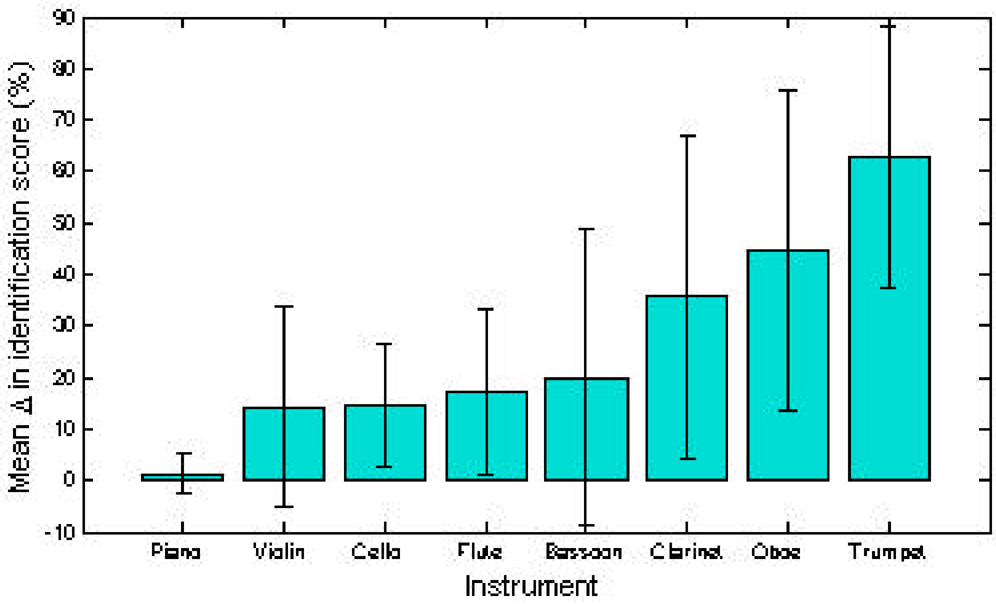


Figure 16: Mean change in identification score from unaltered to processed conditions, across all subjects, for real instruments. Error bars show standard deviation.

Figure 17 shows confusion matrices for unaltered (left panel) and processed (right panel) real instrument sounds. For each instrument presented, a percentage score is given (a blank entry indicates 0%) indicating the frequency it was labeled as each of the other instruments. Responses along the descending diagonal indicate correct responses. The dotted lines enclose confusions between instruments of the same family. In this case, the families are woodwinds, brass (trumpet), strings and percussion (piano).

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**80**	11.3	6.7		2			
	2	4	**82**	11.3	0.7	2			
	3	0.7	9.3	**75.3**	14	0.7			
	5				**99.3**		0.7		
	6			3.3		**96.7**			
	8						**96**	4	
	9	1.3						**98.7**	
	11								**100**

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**60**	10.3	5.7	1.3	4.3	2	13	3.3
	2	2	**37.3**	19	8.7	9	21.3	1.3	1.3
	3	7.7	11.3	**39.7**	33.7	2.7	1.7	3.3	
	5	0.3	5.3	9	**82.3**	1	2		
	6	5.7	20.7	18	7	**33.7**	7	5	3
	8	0.3	5	1.7	0.3	0.3	**81.7**	10.7	
	9	5	0.7	1			9.3	**84**	
	11			0.7	0.7				**98.7**

Figure 17: Confusion matrices for all subjects for unaltered (left panel) and processed (right panel) real instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

For unaltered real instrument sounds, only a small percentage of confusions occur out-of-family (1.25%). Of the lower correct identification scores, the clarinet (3) was mistaken for the flute (5) more than any other instrument. After the sounds have been processed, more confusions occur overall. For these processed sounds, out-of-family confusions increased to 18.5%. Some notable new confusions were judging the oboe (2) as violin (8), and trumpet (6) as oboe (2).

The overall mean for correctly scoring the unaltered real instrument sounds was 91%. For processed sounds, the overall mean identification score was 64.7%. The difference between these groups of scores is highly significant using the Wilcoxon test (p $<<$ 0.00001).

4.1.2 All subjects - synthesized instruments

Figure 18 shows results for each subject on identification of the unaltered and processed synthesized instrument sounds. Processed scores are shown separately for the full- and half-wave rectification conditions.

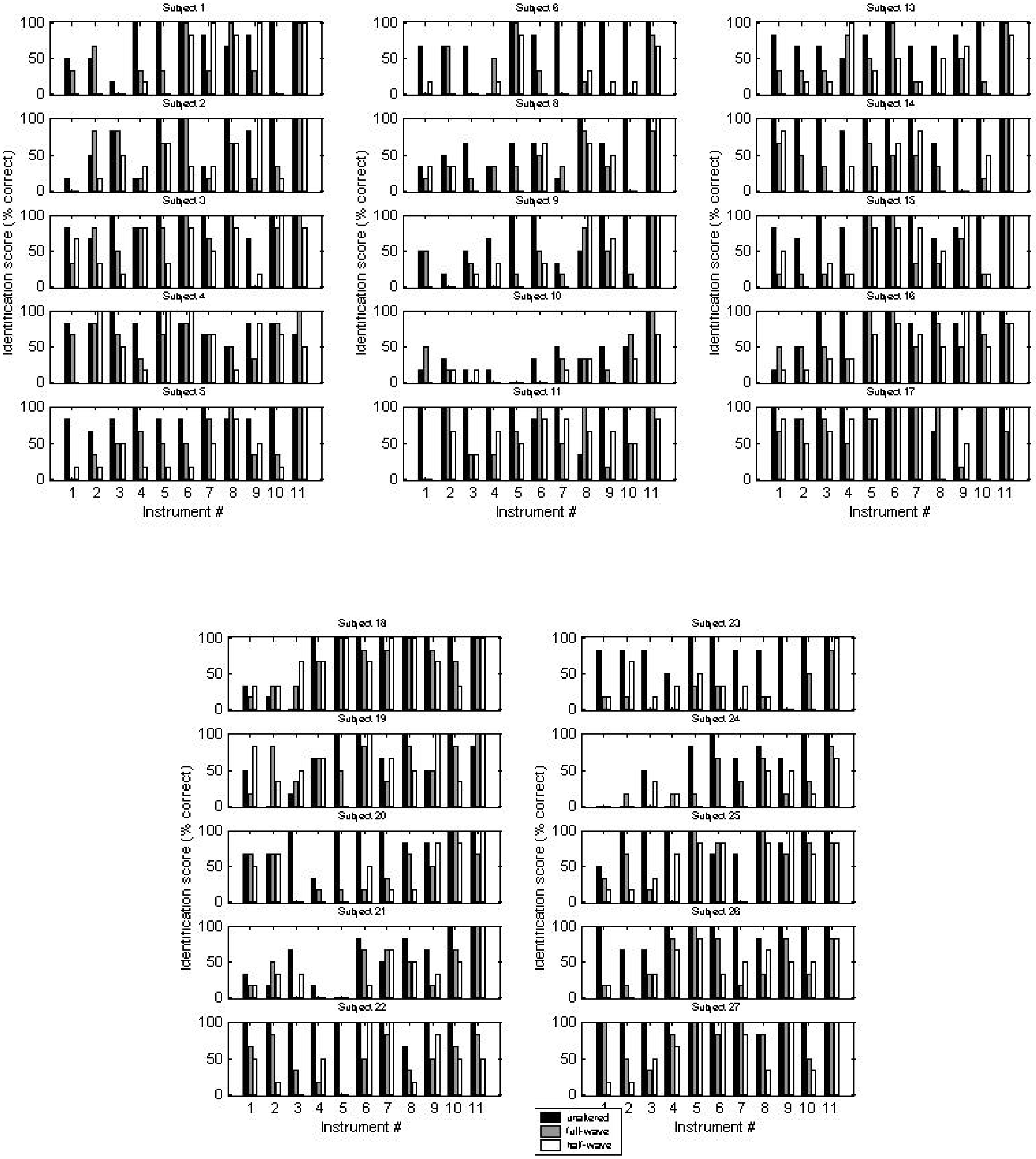


Figure 18: Identification scores for each subject using synthesized instrument sounds. Mean scores for each instrument are shown for unaltered sounds, processed sounds using full-wave rectification, and processed sounds using half-wave rectification.

Again, some subjects seem to be able to identify the unaltered instrument sounds better than others. The average scores and corresponding standard deviation for all subjects for both unaltered and processed synthesized stimuli are shown in Figure 19, for each instrument. The distribution of scores that determine the mean for each instrument is again non-Gaussian, thus the Wilcoxon rank test was used to determine whether the population of scores for the unaltered condition was significantly different that that for the processed condition. For all instruments, the difference in scores between unaltered and processed was significant (p = 0.01).

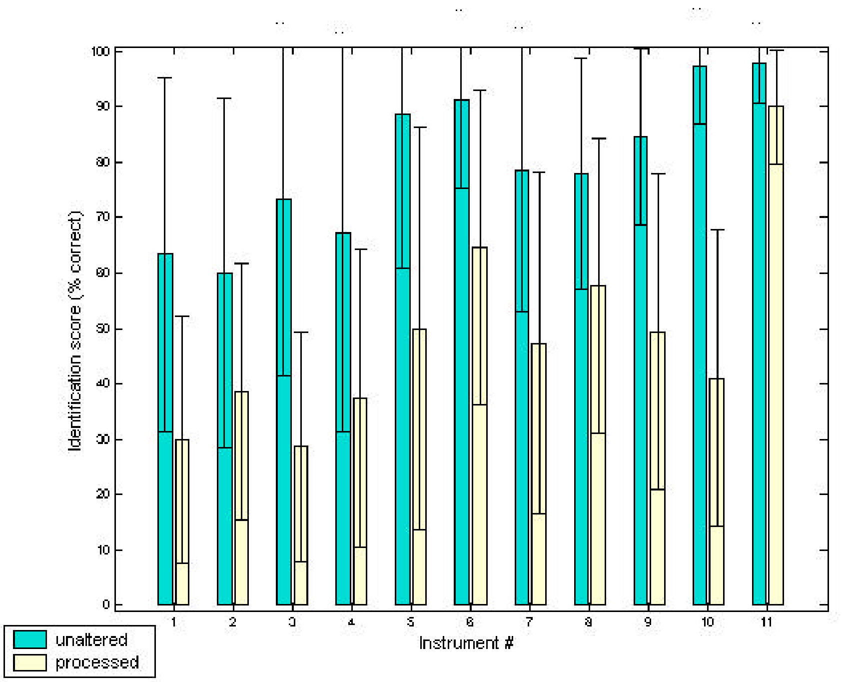


Figure 19: Mean scores across all subjects for each instrument testing unaltered and processed synthesized instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

A comparison of results for full- versus half-wave rectification is shown in Figure 20.

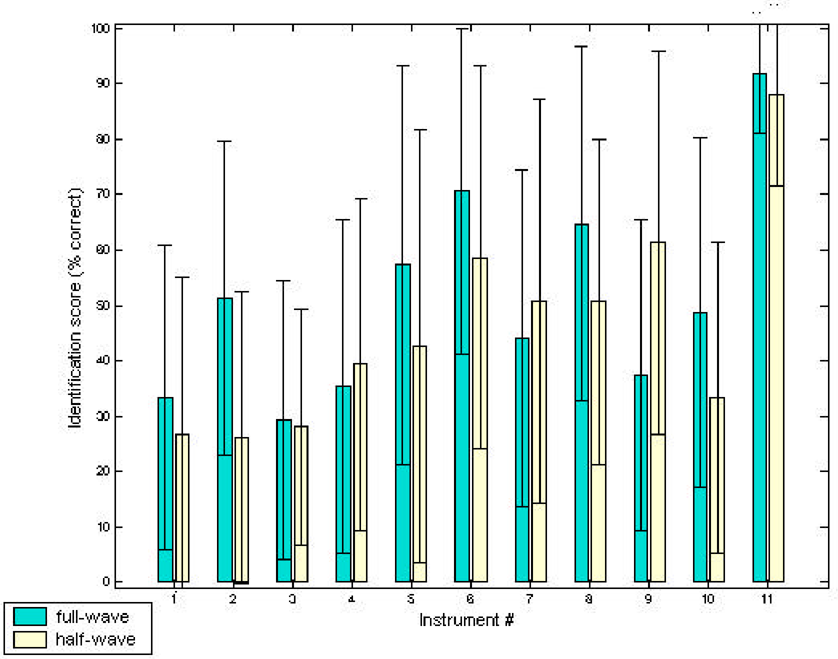


Figure 20: Mean identification scores across all subjects for each instrument on processed synthesized instrument sounds, grouped by full- and half-wave rectification

The full- and half-wave identification scores were significantly different for the oboe (2) (p = 0.01) and the cello (9) (p = 0.05). Full-wave rectification yielded a higher mean score for oboe while half-wave yielded higher identification for the cello.

To demonstrate the ranking of the instruments, the following figure shows the distribution of mean difference scores across all subjects for each of the synthesized instruments.

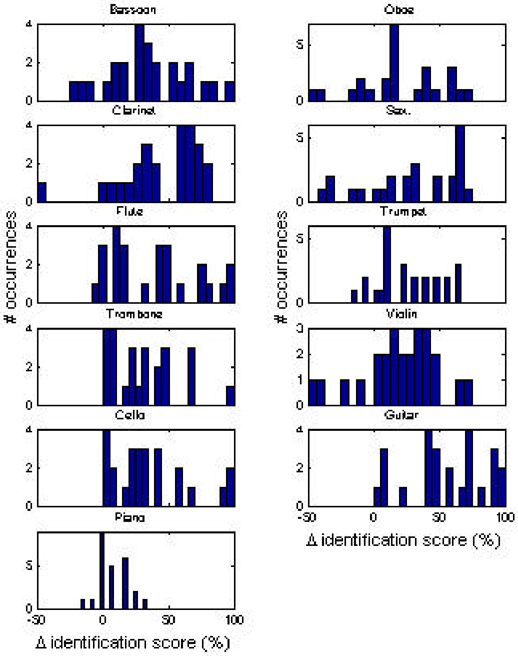


Figure 21: Frequency of occurrence of differences in identification scores between unaltered and processed sounds, across all subjects, for each synthesized instrument.

The mean difference and standard deviation in identification scores for synthesized instruments are shown in Figure 22, which illustrates the increasing order of difficulty in identification (i.e., increased distortion) of each instrument.

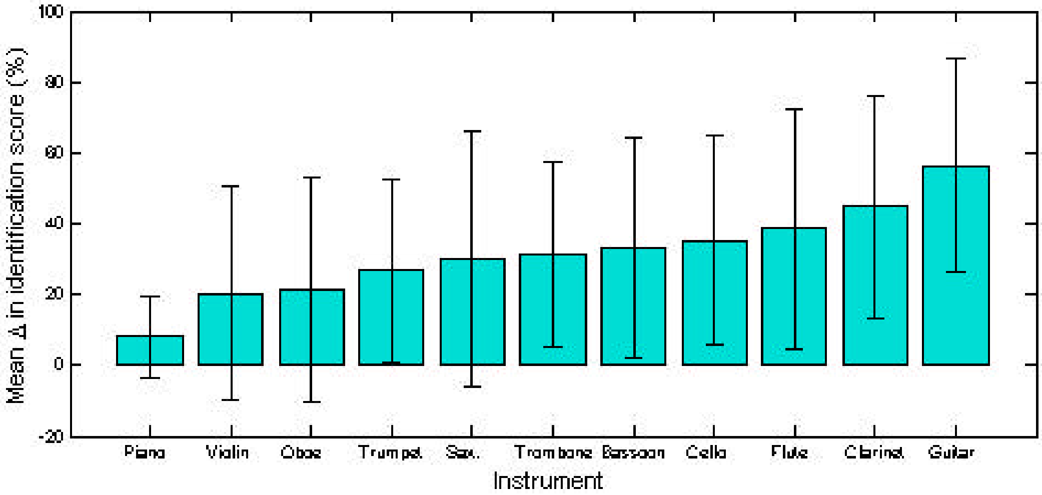


Figure 22: Mean change in identification score from unaltered to processed conditions, across all subjects, for synthesized instruments. Error bars show standard deviation.

Here, the instrument families do not cluster together in terms of difficulty. The means increase very gradually and the deviations are large, indicating that most instruments are similarly affected by processing. The piano and violin remain the easiest to identify, while guitar is the most difficult.

Figure 23 shows the confusion matrices for unaltered and processed cases using synthesized instrument sounds.

		selected										
		1	2	3	4	5	6	7	8	9	10	11
presented	1	**63.3**	16.7	6.7	10.7		0.7	2				
	2	2.7	**60**	20.7	7.3	6	2.7			0.7		
	3	3.3	16.7	**73.3**	3.3	2.7	0.7					
	4	0.7	14.7	8	**67.3**	1.3	6			0.7	1.3	
	5		2.7	3.3	0.7	**88.7**	4.7					
	6		0.7	1.3	1.3	1.3	**91.3**	4				
	7	16	1.3		3.3		0.7	**78.7**				
	8			0.7	1.3		0.7		**78**	19.3		
	9	1.3	1.3						12.7	**84.7**		
	10									1.3	**97.3**	1.3
	11	0.7	0.7		0.7							**98**

		selected										
		1	2	3	4	5	6	7	8	9	10	11
presented	1	**30**	8.7	19	8.3	5.3	2.3	11.3	1.3	9.3	3	1.3
	2	3	**38.7**	16.3	12.7	8.7	6.3	2.3	8.3	1.3	1	1.3
	3	1.7	15.7	**28.7**	15.7	16.7	3	1.7	3.3	0.3	10.3	3
	4	1	19.7	5.3	**37.3**	3	13.3	2.7	2		11.3	4.3
	5	3	5.3	18	3.7	**50**	4.3	0.7	10.7	3.7	0.3	0.3
	6	3	1.7	3.7	9.3	0.3	**64.7**	5.7	0.7	0.7	7.7	2.7
	7	41.3	3.7		2		2	**47.3**		3.7		
	8	1.7	5.3	7.3	4	7	1.3	0.7	**57.7**	15		
	9	5.3	4.3	3.3	5.7	1	1.3	1.3	27.7	**49.3**	0.7	
	10		0.7	1	2	1	0.3	0.7	0.3	0.7	**41**	52.3
	11					0.3		0.3			9.3	**90**

Figure 23: Confusion matrices for all subjects for unaltered (left panel) and processed (right panel) synthesized instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

Out-of-family confusions occur on average 4.97% of the time for the unaltered case. The oboe (2) was mistaken for the clarinet (3) in 20.7% of trials, higher than other instruments in that family. In general, woodwinds show a large number of confusions. Violin (8) was confused with cello (9) in 19.3% of trials, although this confusion is quite common in general. After processing, out-of-family confusions occurred in 29.5% of trials. The frequency of selecting the guitar when the piano was played was 52%. Members of the woodwind family confusion showed a large number of confusions, both within and outside of the family.

The overall mean for correctly identifying unaltered synthesized instrument sounds was 80.06%. For the processed case, the overall mean identification score was 48.6%. The difference between these two groups of scores is highly significant ($p \ll 0.00001$).

We will now look at some subgroups of subjects for analysis.

4.2 Top performers for real, unaltered instrument sounds

Since one important question is whether or not someone who had prior knowledge of a particular instrument could identify it through the implant simulation, we select a sub-group of listeners for analysis who scored well on the unaltered instrument sounds. The selection criterion was that each subject correctly identify each unaltered instrument at least five out of the six times it was presented. This is a reasonable cutoff given the performance of all subjects on unaltered sounds. Ten subjects met this criterion: subjects 1, 4, 11, 14, 15, 16, 17, 20, 22 and 26. Henceforth these subjects will be referred to as the “expert listener group” for real, unaltered instruments, ELG_R.

4.2.1 ELG_R - real instruments

Mean identification scores for each instrument under the unaltered and processed conditions are shown in Figure 24 for ELG_R. Processing significantly lowers scores for all except the flute ($p = 0.05$) and piano ($p =0.01$).

When comparing results for full- and half-wave rectification, the difference in distribution of scores was significant for the violin (8) ($p = 0.01$) and the flute (5) ($p = 0.05$). Full-wave rectification seemed to improve scores for these two instruments. Figure 25 shows average scores for full- and half-wave rectification.

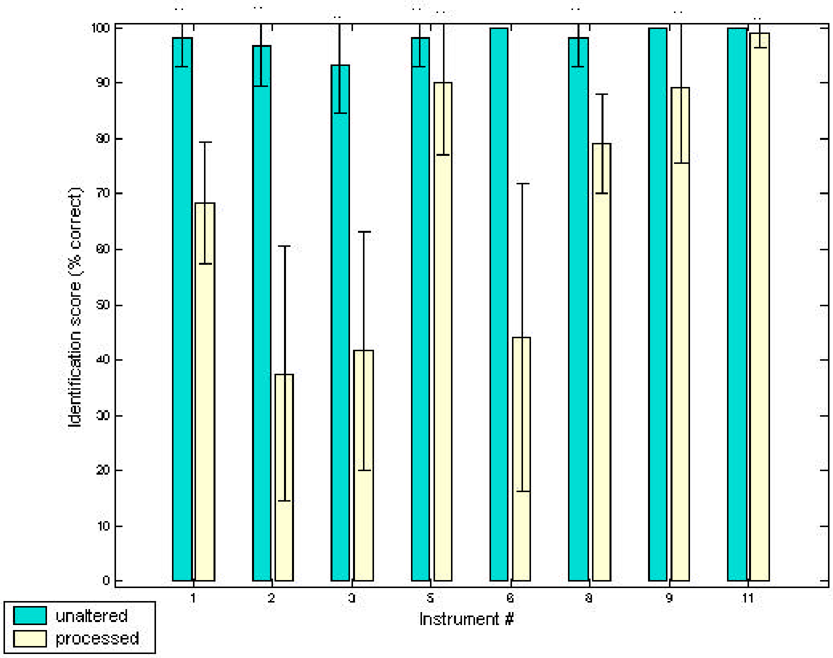


Figure 24: Mean scores for ELG_R for each instrument testing unaltered and processed real instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

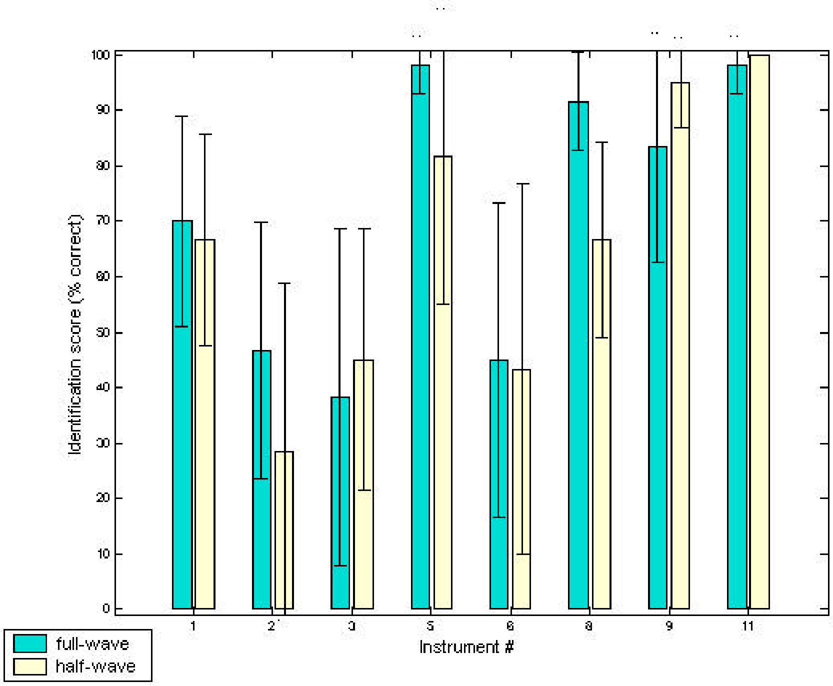


Figure 25: Mean identification scores for ELG_R for each instrument on processed real instrument sounds, grouped by full- and half-wave rectification

Figure 26 shows the distribution of mean difference scores between unaltered and processed conditions for ELG_R on each of the real instruments.

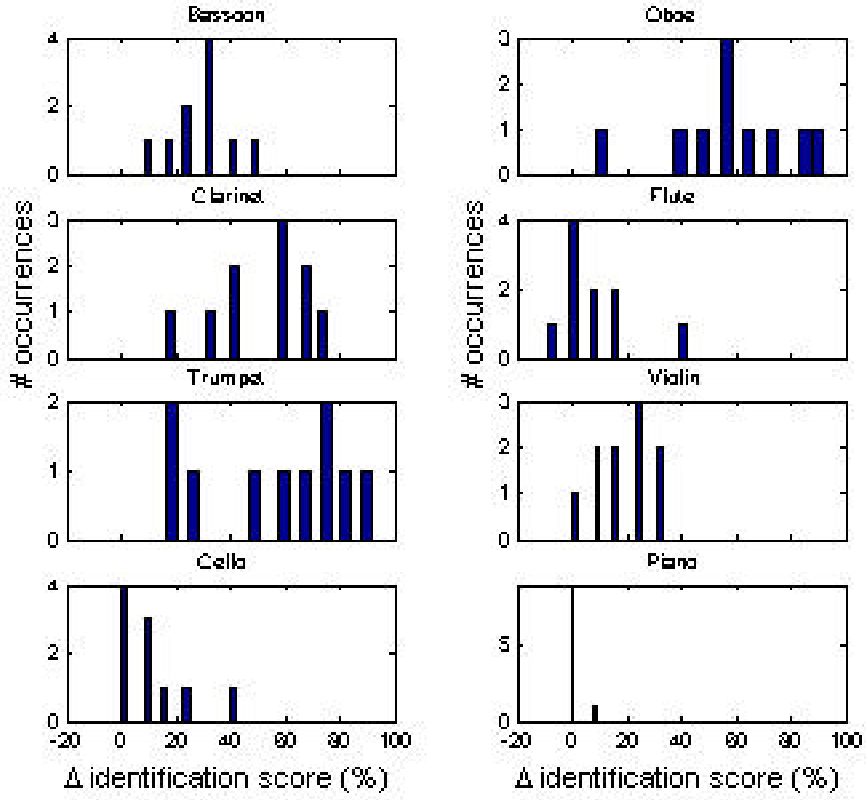


Figure 26: Frequency of occurrence of differences in identification scores between unaltered and processed sounds, for ELG_R, for each real instrument.

The mean difference and standard deviation in identification scores for ELG_R for real instruments are shown in Figure 27, which illustrates the increasing order of difficulty in identification (i.e., increased distortion) of each instrument. Here, the expert subjects could identify the piano most often after processing. Scores for cello and flute are less affected by processing, while the scores for the oboe, trumpet and clarinet are most affected. The ordering here is similar to the ordering presented earlier for all subjects.

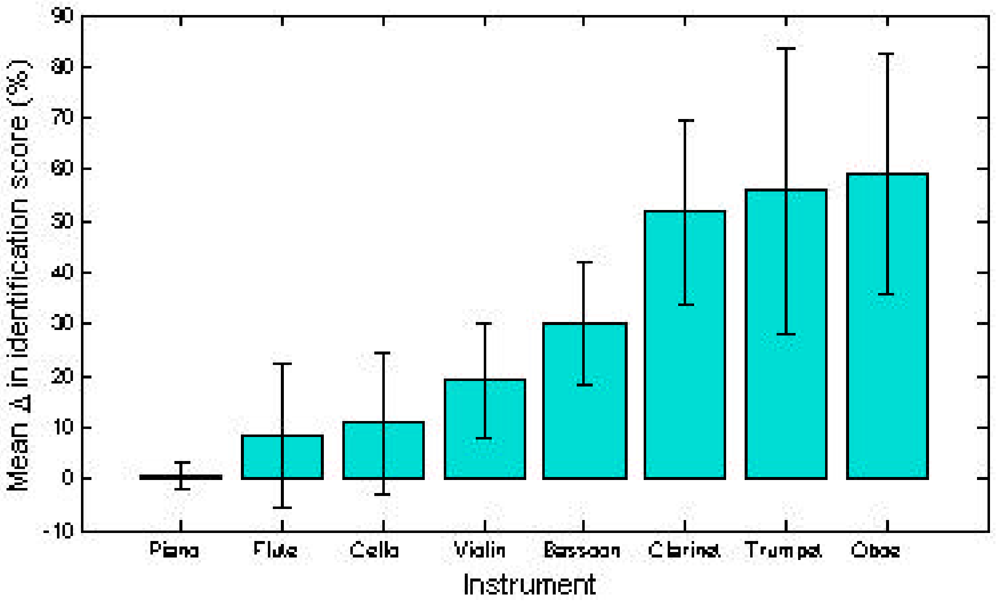


Figure 27: Mean change in identification score from unaltered to processed conditions for ELG$_R$, for real instruments. Error bars show standard deviation.

For unaltered and processed instrument sounds, the confusion matrices with percentage scores are shown in Figure 28.

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**98.3**		1.7					
	2		**96.7**	3.3					
	3		1.7	**93.3**	5				
	5				**98.3**		1.7		
	6					**100**			
	8						**98.3**	1.7	
	9							**100**	
	11								**100**

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**68.3**	5.8	7.5	2.5	0.8	1.7	9.2	4.2
	2	0.8	**37.5**	16.7	10	5	25.8	3.3	0.8
	3	6.7	5	**41.7**	40.8		1.7	4.2	
	5		3.3	3.3	**90**	1.7	1.7		
	6		18.3	14.2	7.5	**44.2**	7.5	5.8	2.5
	8		2.5	2.5	0.8		**79.2**	15	
	9	4.2					6.7	**89.2**	
	11			0.8					**99.2**

Figure 28: Confusion matrices for ELG$_R$ for unaltered (left panel) and processed (right panel) real instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

As expected, since these were the best performers for this instrument class, only a very small percentage of confusions occur out-of-family (0.21%) for the unaltered sounds. However, after sounds were processed, confusions spread outside of families. For these processed sounds, out-of-family confusions averaged 15.8%. The oboe (2) was mistaken for the violin (8) in 25.8% of trials. The clarinet (3) and trumpet (6) also scored rather poorly and were confused with instruments within and outside their families. The clarinet (3) was mistaken for the flute (5) on almost as many trials as it was correctly identified. The trumpet (6) was mistaken for the oboe (2) in 18.3% of trials.

The overall mean for correct identification scores on all unaltered real instrument sounds was 98.1% for ELG$_R$. For the remaining “non-expert” subjects, the overall mean for the same stimuli was 86.3%. The two groups of scores differ significantly ($p = 0.01$), thus the ELG$_R$ group is more capable of identifying these unaltered, real instrument sounds. For processed instruments, the overall mean identification score was 68.6% for ELG$_R$. The non-expert subjects scored 62% for the same task. These scores are not significantly different, thus the information remaining after processing seems equally assessable to experts and non-expert subjects. The difference between the unaltered and processed scores for ELG$_R$ is significant ($p = 0.01$), thus the processing lowered scores overall for this group’s identification of real instruments.

4.2.2 ELG$_R$ - synthesized instruments

Although these 10 subjects were the best performers for the real instruments, they were not the best performers on synthesized, unaltered instrument sounds. However, one can still analyze ELG$_R$'s performance on synthesized instruments.

The average scores for ELG$_R$ for both unaltered and processed synthesized stimuli are shown in Figure 29, along with the error bars corresponding to the standard deviation. The processed scores include scores for both full-wave and half-wave rectification methods.

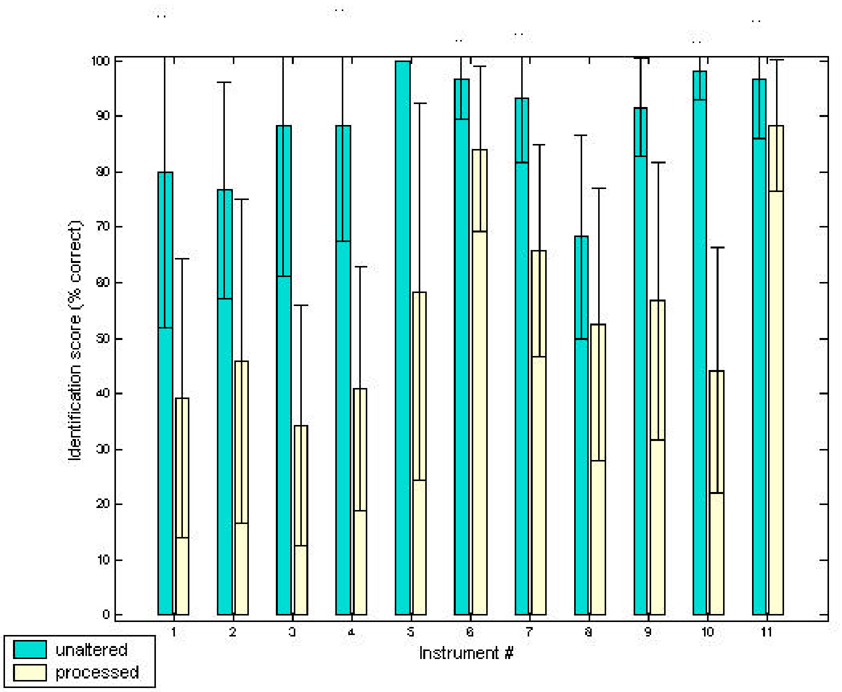


Figure 29: Mean scores for ELG$_R$ for each instrument testing unaltered and processed synthesized instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

When analyzing the differences in score distributions for each instrument from unaltered to processed cases, significant differences occurred for all instruments except the violin ($p = 0.05$). Under stricter significance, scores did not change for the oboe (2), trumpet (6) and piano (11) ($p = 0.01$), in addition to the violin. Thus fewer synthesized instruments were significantly affected by processing when compared to the real instruments for the ELG$_R$.

When comparing results for full- and half-wave rectification, there were significant differences in scores for the oboe (2), trombone (7) and cello (9) ($p = 0.05$). Full-wave rectification yielded higher scores for the oboe only, while half-wave rectification yielded higher scores for the trombone and cello. Figure 30 shows average scores for full- and half-wave rectification.

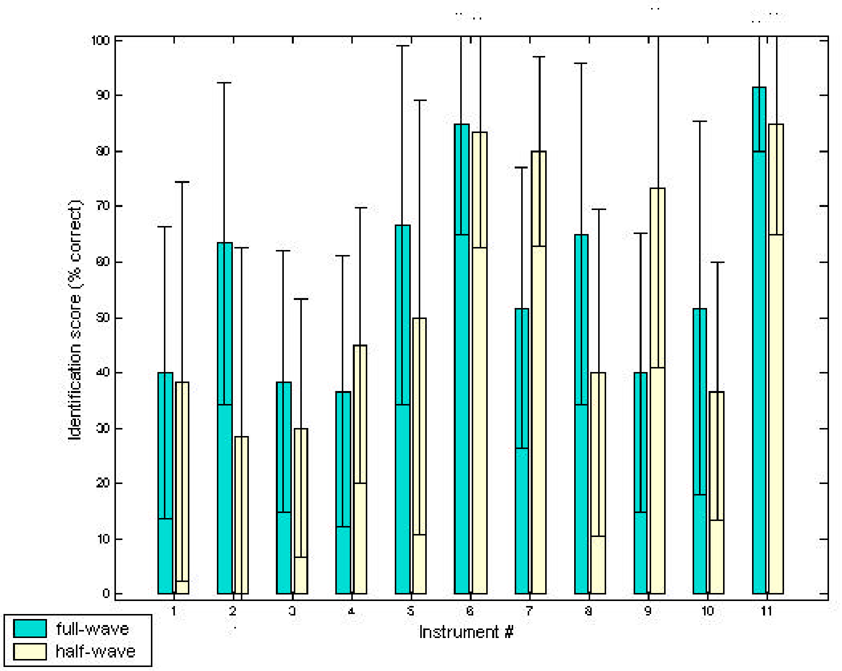


Figure 30: Mean identification scores for ELG$_R$ for each instrument on processed synthesized instrument sounds, grouped by full- and half-wave rectification

Since ELG$_R$ are experts for real instrument sounds, we are not interested in the ranking of synthesized instruments.

For unaltered and processed instrument sounds, the confusion matrices with percentage scores are shown in Figure 31.

		selected										
		1	2	3	4	5	6	7	8	9	10	11
presented	1	**80**	10	5	5							
	2		**76.7**	15	3.3	5						
	3	1.7	10	**88.3**								
	4		10	1.7	**88.3**							
	5					**100**						
	6		1.7				**96.7**	1.7				
	7	6.7						**93.3**				
	8								**68.3**	31.7		
	9		1.7						6.7	**91.7**		
	10									1.7	**98.3**	
	11		1.7		1.7							**96.7**

		selected										
		1	2	3	4	5	6	7	8	9	10	11
presented	1	**39.2**	7.5	16.7	7.5	5	4.2	6.7	0.8	6.7	2.5	3.3
	2	4.2	**45.8**	6.7	17.5	5	10		5.8	0.8	0.8	3.3
	3	0.8	15.8	**34.2**	15	12.5	1.7	1.7	5.8	0.8	9.2	2.5
	4	0.8	18.3	5.8	**40.8**		17.5	0.8	2.5		9.2	4.2
	5	0.8	2.5	22.5	4.2	**58.3**	1.7		5.8	4.2		
	6	0.8	0.8		7.5		**84.2**	1.7	0.8	0.8	2.5	0.8
	7	32.5			1.7			**65.8**				
	8	1.7	2.5	7.5	3.3	11.7			**52.5**	20.8		
	9	8.3	1.7	1.7	6.7	1.7			22.5	**56.7**	0.8	
	10			0.8	0.8		0.8				**44.2**	53.3
	11										11.7	**88.3**

Figure 31: Confusion matrices for ELG$_R$ for unaltered (left panel) and processed (right panel) synthesized instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

For unaltered sounds, out-of-family confusions occur in 1.36% of trials. Of the greater confusions, the violin (8) was mistaken for the cello (9) in 31.7% of trials. After sounds were processed, confusions spread outside of families, averaging 25.1%. Woodwind scores were generally low, and the guitar (10) was mistaken for the piano (11) in 53.3% of trials. The trombone (7) was mistaken for the bassoon (1) in 32.5% of trials. The strings were confused with woodwinds more than any other family.

The overall mean for correctly scoring on all unaltered synthesized instrument sounds for group ELG$_R$ was 88.9%. For processed instruments, the overall mean identification score for this group was 55.5%. These two score distributions differ significantly ($p = 0.01$), illustrating the effect of processing on synthesized sounds.

4.3 Top performers for synthesized, unaltered instrument sounds

Just as we formed a group of subjects who scored highly in the identification of unaltered, real instruments, we selected a group of “expert listeners” for unaltered, synthesized instruments. Although the experiments were slightly different, a direct comparison shows scores for synthesized instruments (80%) are significantly lower overall than those for real instruments (91%) ($p = 0.01$). If we used the same criterion used for selecting “expert listeners” for real instruments, we would only end up with one subject. A more relaxed criterion was applied in this case to select a group of “expert subjects”, herein referred to as ELG_S: each instrument must be correctly identified in at least four out of six trials. Applying this criterion results in eight members of ELG_S: subjects 3, 5, 14, 15 17, 22, 26, and 27.

4.3.1 ELG_S - real instruments

Although our interest lies in how ELG_S performed on synthesized instruments, we proceed with the analysis of results for real instruments first. The average scores for ELG_S for both unaltered and processed real stimuli are shown in Figure 32, along with the error bars corresponding to the standard deviation. The processed scores include scores for both full-wave and half-wave rectification methods.

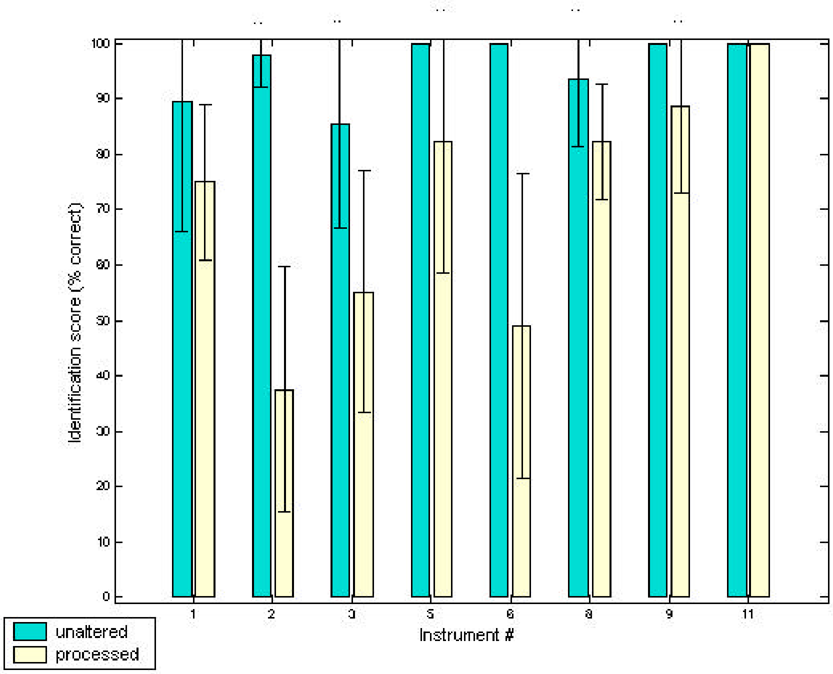


Figure 32: Mean scores for ELG_S for each instrument testing unaltered and processed real instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

The differences in scores between the unaltered and processed cases are significant for the oboe (2) and trumpet (6) ($p = 0.01$), and additionally for the bassoon (1) and clarinet (3) ($p = 0.05$). Thus processing affected only half of the real instruments for ELG_S. Figure 33 shows identification scores for full- and half-wave rectification.

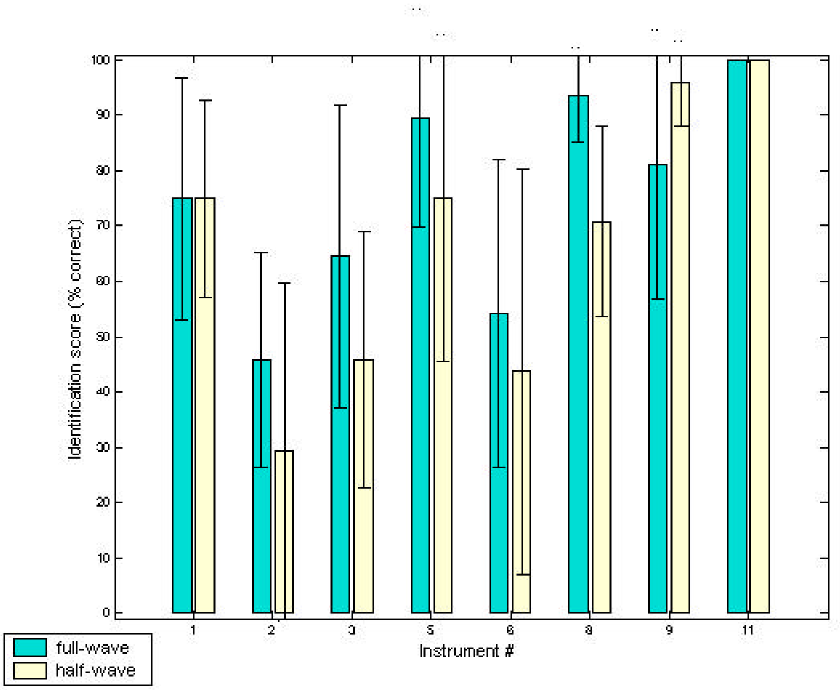


Figure 33: Mean identification scores for ELG_S for each instrument on processed real instrument sounds, grouped by full- and half-wave rectification

Analysis of scores for full- and half-wave rectification found that differences were significant only for violin (8) ($p = 0.05$). Full-wave rectification resulted in a higher mean for the violin.

We are less interested in ranking of real instruments for group ELG_S, since these are experts for synthesized instrument sounds.

For unaltered and processed real instrument sounds, the confusion matrices with percentage scores for ELG_S are shown in Figure 34.

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**89.6**	8.3	2.1					
	2		**97.9**	2.1					
	3		2.1	**85.4**	12.5				
	5				**100**				
	6					**100**			
	8						**93.8**	6.2	
	9							**100**	
	11								**100**

		selected							
		1	**2**	**3**	**5**	**6**	**8**	**9**	**11**
presented	**1**	**75**	5.2	5.2	2.1			9.4	3.1
	2	1	**37.5**	17.7	12.5	7.3	19.8	2.1	2.1
	3	8.3	3.1	**55.2**	24	2.1	2.1	5.2	
	5		2.1	12.5	**82.3**	1	2.1		
	6	2.1	12.5	13.5	9.4	**49**	6.2	6.2	1
	8		3.1	2.1	1	1	**82.3**	10.4	
	9	3.1					8.3	**88.5**	
	11								**100**

Figure 34: Confusion matrices for ELG_S for unaltered (left panel) and processed (right panel) real instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

For unaltered sounds, all confusions occurred within instrument families. The clarinet (3) was mistaken for the flute (5) in 12.5% of trials, while the bassoon (1) was confused with oboe (2) and clarinet (3). After the simulation, confusions spread across families. For these processed sounds, out-of-family confusions averaged 14.7%. The oboe (2) was confused with all other instruments, and 19.8% of these confusions were with the trumpet (8). The trumpet (8) was confused with all instruments as well, with higher confusions for oboe (2) and clarinet (3).

The overall mean for correctly scoring on all unaltered instrument sounds was 95.8%. For processed instruments, the overall mean identification score was 71.2%. The difference these score distributions is significant ($p = 0.01$), illustrating that processing affects ELG_S significantly.

4.3.2 ELG_S - synthesized instruments

By our criterion, ELG_S is the better performing group for unaltered, synthesized instruments. The average scores for ELG_S for both unaltered and processed synthesized stimuli are shown in Figure 35, along with the error bars corresponding to the standard deviation. The processed scores include scores for both full-wave and half-wave rectification methods.

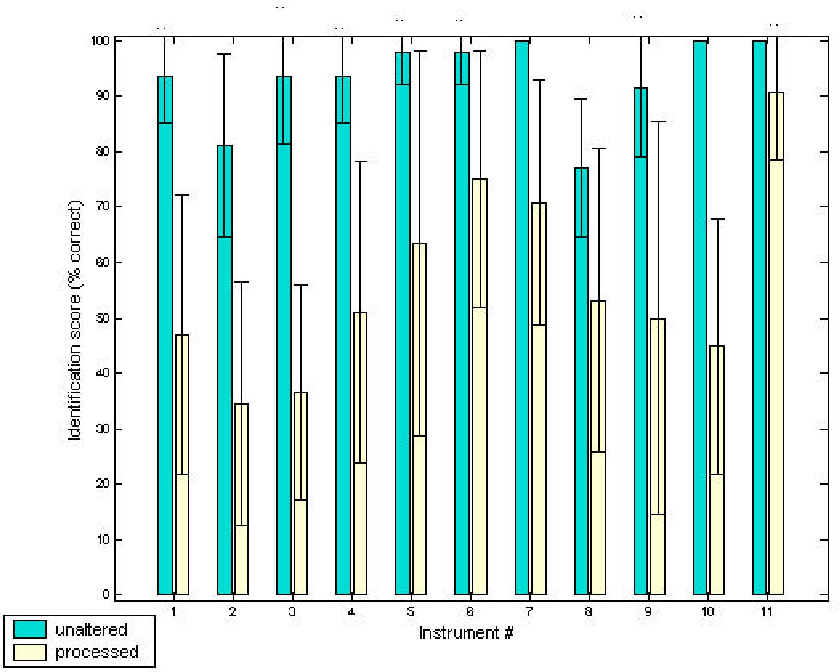


Figure 35: Mean scores for ELG_S for each instrument testing unaltered and processed synthesized instrument sounds (means for processed sounds include full-wave and half-wave rectification scores)

Differences in scores from unaltered to processed sounds occur for all except the violin (8), cello (9) and piano (11) ($p = 0.01$). However, the change is significant for the cello under a more relaxed significance level ($p = 0.05$).

When comparing results for full- and half-wave rectification, the difference in score distributions was not significant except for the oboe (2) ($p = 0.05$), with an improvement using full-wave rectification. Figure 36 shows average scores for ELG_S for full- and half-wave rectification methods for synthesized instruments.

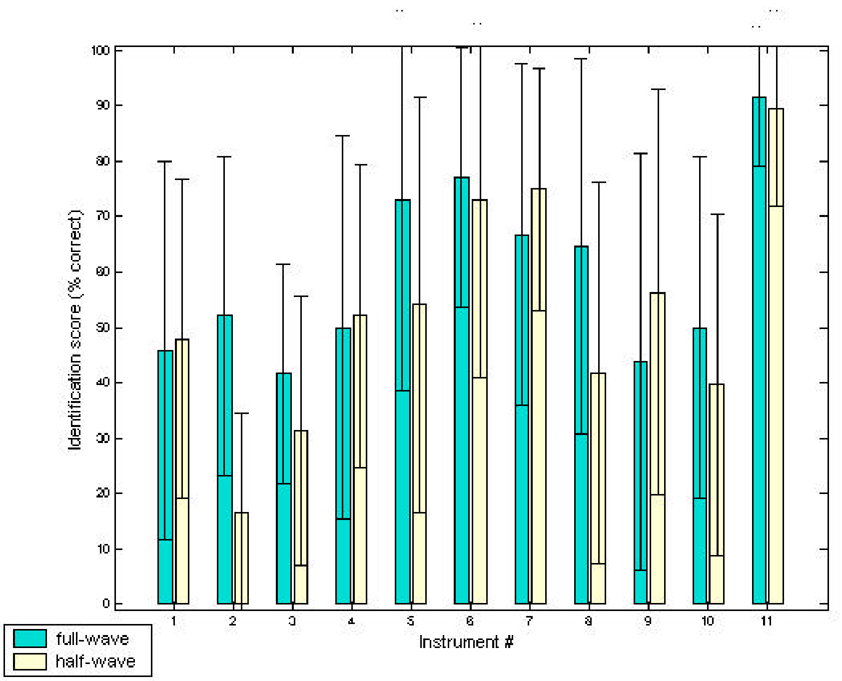


Figure 36: Mean identification scores for ELG$_S$ for each instrument on processed synthesized instrument sounds, grouped by full- and half-wave rectification

To aid in calculating the ranking of the instruments, Figure 37 shows the distribution of mean difference scores for ELG$_S$ on each of the synthesized instruments.

The mean difference and standard deviation in identification scores for synthesized instruments for ELG$_S$ are shown in Figure 38, which illustrates the increasing order of difficulty in identification (i.e., increased distortion) of each instrument.

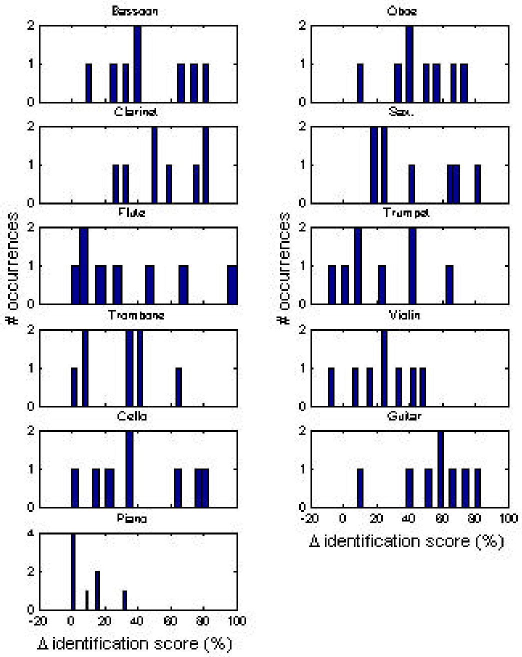


Figure 37: Frequency of occurrence of differences in identification scores between unaltered and processed sounds, for ELG_S, for each synthesized instrument.

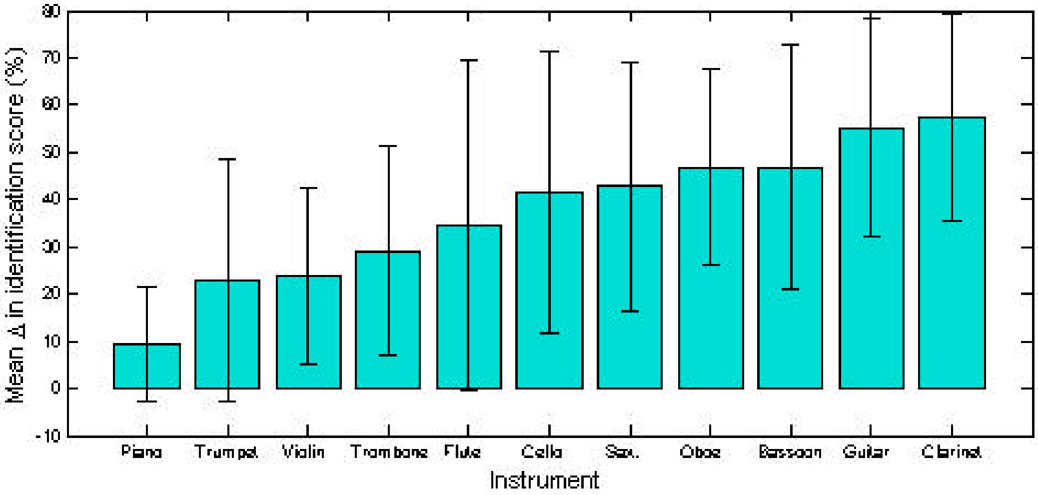


Figure 38: Mean change in identification score from unaltered to processed conditions for ELG_S, for synthesized instruments. Error bars show standard deviation.

Here, the expert subjects could identify the piano most often after processing. The synthesized trumpet seems easier to identify after processing than the acoustic version. Scores for the guitar and clarinet lowered the most after the processing.

For unaltered and processed synthesized instrument sounds, the confusion matrices with percentage scores for ELG_S are shown in Figure 39.

presented	selected										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	**93.8**	4.2	2.1								
2		**81.2**	14.6		4.2						
3		6.2	**93.8**								
4		2.1	2.1	**93.8**		2.1					
5			2.1		**97.9**						
6						**97.9**	2.1				
7							**100**				
8								**77.1**	22.9		
9								8.3	**91.7**		
10										**100**	
11											**100**

presented	selected										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	**46.9**	7.3	16.7	3.1	5.2		7.3	1	9.4	1	2.1
2	2.1	**34.4**	15.6	17.7	7.3	9.4		8.3	1	2.1	2.1
3	1	17.7	**36.5**	18.8	9.4	3.1	1	5.2	1	5.2	1
4		15.6	3.1	**51**		13.5		3.1		10.4	3.1
5	1	2.1	20.8	1	**63.5**	2.1	1	4.2	4.2		
6	1	1	1	10.4		**75**	1	1	1	7.3	1
7	24					2.1	**70.8**		3.1		
8	1	3.1	7.3	2.1	14.6			**53.1**	18.8		
9	7.3	1	2.1	7.3	1			31.3	**50**		
10			1			1				**44.8**	53.1
11										9.4	**90.6**

Figure 39: Confusion matrices for ELG_S for unaltered (left panel) and processed (right panel) synthesized instruments. Dotted lines show grouping of instrument families. Numbers show percent mean identification score.

Out-of-family confusions occurred in only 0.19% of trials for unaltered synthesized sounds, which is expected since these are the "expert listeners" for synthesized instruments. The violin (8) was mistaken for the cello (9) in 22.9% of trials. After sounds were processed, out-of-family confusions averaged 24.1%. Woodwind scores were generally low, and the guitar was mistaken for the piano in 53.1% of trials (but the piano is only mistaken for the guitar in 9.4% of trials). The clarinet (3) was confused with all instruments.

The overall mean for correctly identifying all unaltered synthesized instrument sounds for group ELG_S was 93.4%. In comparison, the remaining "non-expert" listeners for unaltered synthesized sounds correctly identified 73.8% of sound stimuli. This "select" group for synthesized instruments did perform significantly better than the non-expert listeners ($p = 0.01$) for unaltered sounds. For processed instrument sounds, the overall mean identification score for ELG_S was 56.1%, compared with 45.1% for the remaining non-expert listeners. The distribution of these scores differs significantly ($p = 0.01$), thus ELG_S and non-expert listeners are not using the same amount of information available after processing for identification. For ELG_S, the difference in scores from unaltered (93.4%) to processed (56.1%) is highly significant ($p \ll 0.00001$). Therefore, although these are "expert" subjects for unaltered sounds, processing affects their identification scores significantly. The drop in score from unaltered (73.8%) to processed (45.1%) for the non-experts was also highly significant ($p << 0.00001$).

It is interesting to see whether the performance of groups ELG_R and ELG_S is similar on identification of processed instruments. In comparing scores for correctly identifying real, processed sounds (mean 68.6% for ELG_R, 71.2% for ELG_S), these differences were not found to be significant. For synthesized, processed sounds (mean 55.5% for ELG_R, 56.1% for ELG_S), the difference in score distributions is also not significant. Thus, choosing "expert" subjects in one instrument class does not imply that scores will be better for this group for processed sounds in that same class, compared with "expert" subjects from another class.

4.4 Effect of Experiment Order

Since half of the subjects tested on real instruments first, while the other half began with synthesized instruments, we now determine if this had any effect on the scores achieved. Those who tested on real instruments first will be referred to as group R1, while those who listened to synthesized instruments first will be group S1. Recall that subjects listened to processed followed by unaltered sounds during the test. Thus, we hypothesize that group R1 might perform better than S1 on unaltered and processed synthesized instruments, since they will have had the opportunity to become familiar with the test using real stimuli. Similarly, we might expect group S1 to perform better than R1 on unaltered and processed real instrument sounds, since they will have become familiar with the test from listening to synthesized stimuli. We are not interested in the complete analysis of the effects of processing on each instrument, nor in the differences between full- and half-wave rectification methods. We only aim to analyze overall scores.

4.4.1 Group R1

4.4.1.1 Real instruments

Over all instruments, group R1 scored 92.1% correct for unaltered sounds and 65.5% correct on processed sounds. The difference in the groups of scores is significant (p = 0.01).

4.4.1.2 Synthesized instruments

Over all unaltered, synthesized instrument sounds, group R1 scored 83% correct. After processing from the simulation, the mean was 50.6%. Thus for processed sounds, the drop in scores from the unaltered case is significant (p = 0.01).

4.4.2 Group S1

4.4.2.1 Real instruments

Over all instruments, group S1 scored 89.8% correct for real, unaltered sounds and 63.8% for processed sounds. The difference between the two groups of scores is significant (p = 0.01).

4.4.2.2 Synthesized instruments

Over all instruments, group S1 scored 76.9% correct for unaltered synthesized instrument sounds and 46.5% for processed sounds. The drop in scores from the unaltered case is significant (p = 0.01).

4.4.3 Comparison of R1 and S1

We might expect that for unaltered, real instrument sounds, group S1 (mean 89.8%) would perform better than R1 (mean 92.1%) if a learning effect was present. The difference between the scores for the two groups is not significant. Thus, no effects of testing order were exhibited. For scores on unaltered, synthesized instrument sounds, we would expect group R1 (mean 83%) to score higher than S1 (76.9%) if there was a learning effect. The difference between the two groups is not significant thus no learning effect occurred for either group. These findings present a strong argument against the learning effect.

For processed, real instrument sounds, a learning effect would be evident if group S1 scored higher than group R1. The difference in scores between S1 (mean 63.8%) and R1 (mean 65.5%) is not significant. For processed, synthesized instrument sounds, a learning effect would be evident if group R1 scored higher than S1. In this case, the difference in scores between R1 (mean 50.6%) and S1 (mean 46.5%) is not significant. Thus, no learning effects were observed, despite the possible difference in level of difficulty of identification of real or synthesized instrument sounds.

4.5 Summary of Results

The following table summarizes mean scores for all instruments pooled together for each group of subjects analyzed, for unaltered and processed sounds. Note that the significance values (**) refer to comparisons between the unaltered and processed means (for example, the mean 68.6% for ELG$_R$ on real, processed instruments is significantly different from 98.1% for real, unaltered instrument sounds).

Table 6: Average scores for all groups on each experiment identification task

**: p = 0.01 (between unaltered and processed)	REAL, unaltered	REAL, processed	SYNTHESIZED, unaltered	SYNTHESIZED, processed
ALL subjects	91%	64.7%**	80.1%	48.6%**
ELG$_R$	98.1%	68.6%**	88.9%	55.5%**
ELG$_S$	95.8%	71.2%**	93.4%	56.1%**
R1	92.1%	65.5%**	83%	50.6%**
S1	89.8%	63.8%**	76.9%	46.5%**

Table 7 shows the ranking order of the instruments, in order of increasing difference in score between the unaltered and processed cases.

Table 7: Ranking of differences between scores of unaltered and processed sounds

Subjects	Instrument class	Instrument ranking (least affected → most affected by processing)
All	real	piano (11) → violin (8) → cello (9) → flute (5) → bassoon (1) → clarinet (3) → oboe (2) → trumpet (6)
All	synthesized	piano (11) → violin (8) → oboe (2) → trumpet (6) → saxophone (4) → trombone (7) → bassoon (1) → cello (9) → flute (5) → clarinet (3) → guitar (10)
ELG$_R$	real	piano (11) → flute (5) → cello (9) → violin (8) → bassoon (1) → clarinet (3) →trumpet (6) → oboe (2)
ELG$_S$	synthesized	piano (11) → trumpet (6) → violin (8) → trombone (7) → flute (5) → cello (9) → saxophone (4) → oboe (2) → bassoon (1) → guitar (10) → clarinet (3)

The following table summarizes the mean percentages for within-family and out-of-family confusions.

Table 8: Confusions for within / out-of-family confusions for each identification condition

	REAL, unaltered	REAL, processed	SYNTHESIZED, unaltered	SYNTHESIZED, processed
ALL subjects	98.8% / 1.25%	81.5% / 18.5%	95% / 4.97%	70.5% / 29.5%
ELG$_R$	99.8% / 0.21%	84.2% / 15.8%	98.6% / 1.36%	74.9% / 25.1%
ELG$_S$	100% / 0%	85.3% / 14.7%	99.8% / 0.19%	75.9% / 24.1%

4.6 Subjective Comments

In a questionnaire designed for the experiment, subjects commented on their experience during the testing: ease of the identification task and instruments that may have been easier or more difficult to identify. In some of the responses it was not clear if the subject was referring to real or synthesized instruments, but some trends were noted. Thirteen subjects listed the piano as either "easy" or "easiest" to identify. Five mentioned that the experiment with synthesized instruments was more difficult. Many suggested they identified sounds based on characteristic features, such as pitch range, non-harmonic characteristics (bowings, "breathy" attacks, vibrato, length of decay). One subject thought that processed instrument sounds could be learned over time, which (if true) is an important point for CI listeners. Six subjects mentioned explicitly that the processed sounds were "difficult" or "very difficult" or "impossible". Oddly enough, one subject mentioned that the processed sounds resembled Bartok's compositions.

4.7 Discussion

We will examine how our results indicate that cochlear implant processing affects instrument identification, and compare our findings with relevant studies in the literature.

4.7.1 Real instruments

All subjects were able to correctly identify real, unaltered instrument sounds in 91% of trials. This figure is higher than the 69% reported by Martin [Martin, 1999] for 10-sec phrases. This might be due to the fact that the closed-set paradigm in this experiment did not include any choices that were not played to the subjects. In addition, the number of instrument stimuli (and thus choices) for our experiment was either eight or 11, while in Martin's experiment, 19 stimuli were heard (with 27 possible choices). Our data is closer to Kendall's results (74.2% non-music majors, 94.6% music majors for phrases) [Kendall, 1986], but Kendall used stimuli from only three instruments. In our experiment, it is not surprising that very few confusions between instruments occurred outside of families, since subjects claimed to be familiar with the instruments beforehand. Martin's musically-trained subjects also identified the majority of instruments within the correct family (96.9%).

One interesting and important phenomenon arose from the experiment with unaltered instrument sounds. Our results show that self-assessment of musical ability is probably not sufficient for selecting a group of subjects for a particular task. In our experiment, there was considerable variability in the subjects' ability to identify real, unaltered instrument sounds even though they may have indicated as having several years of formal music training in their self-evaluation. Figure 11 illustrated these discrepancies between self-assessment and identification scores for all instruments. It was for this reason that we selected "expert listeners" for real and synthesized instruments to properly examine the effects of processing on instrument identification.

Processing for real instruments lowered scores for all except the piano, with an average identification rate of 64.7% correct for all subjects. This figure was significantly different from the 91% average for unaltered sounds. The mean for processed scores is slightly higher than other measures made on implanted listeners: Fujita and Ito's average of 56% correct [Fujita and Ito, 1999], and the 46.55% reported by Gfeller [Gfeller et al., 2002] (recall that subjects heard eight instruments in Gfeller's test but could select from 16, making the task slightly more difficult). Processing increased the confusions across families, thus indicating a degradation in sound. Oboe was often mistaken as violin, and trumpet as oboe.

While full-wave rectification improved recognition for flute and violin and half-wave improved recognition for the cello, overall results for full- versus half-wave rectification were quite similar.

When examining the extent of the effect of processing, the difference in mean scores for unaltered and processed were rank-ordered to determine which instruments lowered in score to a higher degree. For real instruments, the piano was easiest to identify after processing, followed by the bowed strings. The woodwinds clustered together at a middle-level of difficulty, while the trumpet lowered in score the most. We saw from Chapter 3 that the trumpet's temporal envelope is distorted, and that harmonics were not evenly-spaced in the processed sounds. For half-wave rectification, extra low-frequency content was added. These factors could have contributed to the lowering of the score.

4.7.2 Synthesized instruments

For unaltered sounds, subjects scored 80.1% correct on average. Only one study with implanted listeners has used synthesized instrument stimuli, played from an electronic keyboard [Fujita and Ito, 1999]. These subjects correctly identified 56% of five synthesized instruments, after training on the sounds. Despite the differences in stimuli and number of instruments for our experiment, a direct comparison of unaltered real and synthesized sounds showed that the difference was significant (mean scores for synthesized sounds were lower). This agrees with some of the subjects' comments that the synthesized instruments were more difficult to identify. This could also be due to the fact that people are less familiar with synthesized instrument sounds, and some sounds are not as natural-sounding as their acoustic counterparts.

Processing lowered scores to a mean of 48.6% for all subjects, and this group of scores was significantly different from those for unaltered sounds ($p = 0.01$). The out-of-family confusions increased from 5% to almost 30% when processed, illustrating the distorting effect of cochlear implant processing. The guitar was often confused with the piano. This might not be surprising given that the instrument acoustics are similar – guitar sounds are created by plucking a stretched string, while piano strings are hammered, but both resonate over a wooden soundboard. However, interestingly this confusion did not occur in reverse. Perhaps this is due to subjects' familiarity with the piano, whereas the synthesized guitar was less familiar and thus judged as piano. Woodwind confusions were also prevalent, both within and outside of the woodwind family. Processing causes the "peaks" in confusions to shift to new confusions, thus indicating distortions.

The difference between full- and half-wave rectification was only significant for the oboe and cello. As was the case with real instruments, half-wave rectification caused higher scores for the cello. Thus one could recommend using half-wave rectification when listening to the cello, since it seemed to improve scores for real and synthesized instruments, but this would not be a recommendation that is robust for all instruments. It would be interesting to assess the qualitative differences between full- and half-wave rectification, to see if one makes the sounds less harsh or more pleasant.

For synthesized sounds, the piano is easiest to identify after processing, as was the case for real instruments. The violin also ranked in a similar place as for real sounds. However, the woodwinds are spread out in the middle along with the trombone. The identification scores for the guitar are lowered most from the processing. From Chapter 3, we saw that the spectrum of the guitar is very different after processing; there is a large amount of added high-frequency content. This may be a factor in the lowering of the scores.

4.7.3 Expert listeners

The groups ELG_R and ELG_S did not differ significantly in their identification of processed instrument sounds, for both real and synthesized instruments. Considering each group compared to the non-experts, processing for real instruments affected ELG_R and the remaining non-experts equally. However, processing for synthesized instruments affected ELG_S more than the non-experts. Most importantly, the processing did significantly affect scores for both real and synthesized instruments for both ELG_R and ELG_S. Thus, these "experts" who could otherwise identify unaltered instrument sounds showed lower performance scores once sounds were processed by the simulation. This illustrates the effect of processing most evidently.

For ELG_R, real instruments most affected by processing were the trumpet and the oboe. Observing the time and frequency characteristics for the oboe in Chapter 3, it doesn't seem (from a visual perspective) that the oboe appears most affected by processing; thus one must exert caution in drawing conclusions from the visual representations of the signals alone.

For ELG_S, synthesized instruments most affected were guitar and clarinet.

For both groups and instrument classes, the piano remained easiest to identify after processing.

4.7.4 Testing order

The order in which the test conditions were presented (real versus synthesized) did not significantly scores for processed sounds (which are the sounds that subjects heard first during the experiment), since groups R1 and S1 were affected equally. This is favorable since it seems no bias was introduced by ordering effects.

4.8 Summary

Our main objective was to determine to what extent the processing of a cochlear implant distorts or limits the information needed to identify musical instruments. By examining results from our expert listeners, who should otherwise be able to identify unaltered instruments proficiently by our criterion, we found that processing significantly decreases identification scores. This applies to both real and synthesized instrument sounds. Confusions between all instrument families increase for processed sounds and are not as confined to within-family confusions as in the unaltered case. Since scores obtained from this experiment were comparable to those obtained from implanted listeners, this indicates that both groups are using nearly as much information in the processed sounds as possible. This also emphasizes that the problem lies in the processing of the implant and may not exclusively be a result of physiological limitations of implant listeners.

Some instruments are more affected by processing than others. Identification scores for the piano, for all groups of subjects and for both real and synthesized conditions, changed the least from unaltered to processed cases. This may be because the piano is a very common instrument with which many people are familiar. This may also be due to the fact that it was the only percussive string instrument of the test stimuli. For real instrument sounds, scores for the oboe and trumpet lowered the most after processing. For synthesized sounds, the guitar scores decreased the most. This seems surprising given that it is the instrument most similar to the piano. However, since many guitar-piano confusions occurred, it is likely that this brought down the average score. In contrast, the clarinet, for which scores also decreased by a large mount, was confused with all other instruments.

The difference between full- and half-wave rectification for real instruments was only significant for the violin and flute for ELG_R. Full-wave rectification yielded higher recognition scores for these instruments. For synthesized instruments, ELG_S correctly identified the oboe with a significantly higher frequency using full-wave rectification. Overall, the effect of rectification does not seem to be large and consistent across instruments. Because music is often composed for several instruments playing simultaneously, it would not be wise to recommend one rectification method over another.

In the design of the experiment, we debated using real versus synthesized instrument sounds. Because the synthesized instrument sounds were more controlled we hoped they would give us a better indication of the effect of processing. Visual inspection of the distortions produced by the simulation did not provide insight that accounted for the relative difficulty that subjects experienced when listening to the processed stimuli. A more formal, quantitative exploration of this relationship should be considered for future work in this area. In addition, the differences in scores between real and synthesized instruments could be compared more accurately if the experiments were identical.

Identification scores for processed real instruments were not significantly different for ELG_R and non-experts, indicating these groups were using the same amount of information available for identification. However, processing for synthesized instruments resulted in different scores for ELG_S and non-experts. In this case, experts and non-experts were not using the same amount of information available for identification of the processed sounds. Thus one cannot generalize that experts perform better than non-experts on processed sounds for both instrument classes.

Whether one tests real instruments or synthesized instruments first does not affect performance. Groups R1 and S1 were shown to perform equally on the testing condition for which they might have been less prepared due to testing order.

This summarizes the main findings from our experiments. The next chapter will present some concluding remarks and outline areas for further research.

CHAPTER 5 CONCLUSION

The goal of this thesis research was to examine the extent to which the sound processing of a cochlear implant affects identification of musical instruments by musically-trained listeners. The motivation for studying this phenomenon stemmed from the studies that have been done on recipients of cochlear implants, which showed that cochlear-implanted listeners score lower than their normal-hearing counterparts on melodic and timbral identification tasks. While the lack of recognition or identification of instruments does not necessarily imply that the music-listening experience is not pleasurable, it is important to know whether musical instruments are being represented as accurately as possible so that future design iterations may be modified to account for this phenomenon.

The cochlear implant is designed to mimic the behavior of a healthy cochlea, taking advantage of the tonotopic organization of the basilar membrane (discussed in Chapter 2). It conveys sound by translating incoming acoustic sounds into electrical stimuli, by means of extracting temporal envelope information from the outputs of bandpass filters. The information extracted is sufficient for top implant recipient performers to communicate without lip-reading (see Loizou [Loizou, 1999] for a comparison of scores for various processing schemes). However, the information is lacking for conveying musical instrument sounds. We observed how the processing adds and/or removes spectral information from the different instrument sounds: in several cases the temporal envelope structure is preserved, but for some instruments the envelope is significantly distorted.

To motivate our research, we compared studies of instrument identification for normal-hearing and implanted listeners. Previous studies with normal-hearing listeners performing identification of unaltered instrument sounds showed that musically-trained listeners scored between 67% [Martin, 1999] and 95% correct [Kendall, 1986] for musical phrases, depending on the details of the experimental protocol (e.g., the number of response alternatives). However, for implanted listeners, studies on instrument identification reported scores ranging from 13.5% (for identification of nineinstruments [Gfeller and Lansing, 1991]) to 56% (for identification of five instruments [Fujita and Ito, 1999]). In addition, subjects reported that some instruments sounded unpleasant and that enjoyment of certain musical genres had declined post-implantation. In addition to instrument identification, several other music-related phenomena were examined (pitch, melody and rhythm perception), since one must understand what factors are influencing perception of musical instruments. Pitch perception in implanted subjects seems to depend on location and rate of direct electrical stimulation. Discrimination of intervals has the resolution of a semi-tone, only over a small range of frequencies. For tests involving recognition of short melodies, a range of 39% [Fujita and Ito, 1999] to 78% [Gfeller, 1991] correct was obtained, depending whether the test was open- or closed-set, and whether or not the subjects were familiar with the melodies prior to the experiment. Armed with these statistics, and with the knowledge of factors that influence performance in instrument identification (details of the test setup, characteristics of the particular stimuli, and musical training of subjects), we designed an experiment to test identification using a simulation of the sound processing strategy commonly used in current cochlear implant systems.

Our experiment tested the ability of 25 subjects to identify short musical phrases peformed by eight real or 11 synthesized musical instruments. Expert listeners were chosen for each instrument class to reduce the likelihood that performance depended on familiarity with the instrument sounds. Expert subjects for real instruments (ELG_R) achieved a mean identification score of 98.1% correct for unaltered instrument sounds and 68.6% correct for processed sounds. The difference between these two groups of scores is significant. Processing resulted in lowered scores for each instrument individually, with the exception of the piano and flute. In comparing rectification methods, full-wave rectification yielded higher recognition scores for flute and violin. It does not seem appropriate to recommend one rectification method over the other, since music is often heard with many instruments playing simultaneously. Instruments most affected by processing were the oboe and trumpet.

The scores for our subjects for processed sounds are slightly higher than the scores reported by Gfeller (13.5% and 47% for two different studies) and Fujita and Ito (56%). It could be that subjects who participated in our experiment had more musical training, and that by not simulating hearing loss, our normal-hearing subjects could use more of the degraded information to identify the sounds than cochlear-implanted subjects. The test protocols and stimuli differed among all experiments, which also contributes to the variability. In addition, the simulation we designed was not entirely complete. Automatic gain control and non-linear mapping were not implemented; the effects of these factors could also contribute to the differences. However, the fact that the scores are similar for normal and impaired listeners raises the possibility that implanted listeners are able to extract nearly all the information delivered by their implants.

For synthesized instrument sounds, the average score for expert subjects (ELG_S) scored 93.4% correct on unaltered sounds and 56.1% for processed sounds (a significant difference). These scores together with subjects' comments indicate that the synthesized stimuli were more difficult to identify than the real sounds, though the experiments differed slightly in type and number of sound stimuli. Instruments most affected by processing were guitar and clarinet. When comparing rectification methods using synthesized stimuli, full-wave rectification improved recognition for the oboe for expert subjects. As with the acoustic stimuli, we do not recommend one method over the other.

The results for both real and synthesized instrument stimuli demonstrate that cochlear implant processing substantially reduces the information available for instrument identification. The results of our simulation provide an "upper bound" to the performance potential of implanted listeners using a common current sound processing strategy. The poor peformance represented by this "upper bound" emphasizes the need for research directed at processors that better convey information required for identification and represent of timbre more accurately.

5.1 Future Work

This thesis research highlighted the inability of the processing of a cochlear implant to convey the information needed to identify musical instrument sounds. We now propose several approaches that may lead to an improved representation of this information. Experimentation with combinations of current implant parameters could lead to several applications: a "music setting" that would be customized and optimally-suited for implanted listeners, assistive devices or hearing aids. These parameters include an extension of the cutoff frequencies of the lowest- and highest-frequency bandpass filters, to convey fundamental frequency and upper harmonics, respectively. A steeper-sloped lowpass filter with a lower cutoff frequency would result in envelopes without the high-frequency fluctuations, as indicated by some informal listening tests done by the author. An even better method to extract the envelopes is the Hilbert transform technique. The current rectification methods produce high-frequency energy that is aliased into the lower frequency analysis channels, which could be avoided by using the Hilbert transform. The low sampling rate of 13kHz used to generate the signals is generally not high enough to represent full-bandwidth sound of an orchestra. This rate is more suited to speech signals; thus a higher sampling rate would be desirable to represent the higher frequencies that are present in the harmonics of musical instruments. Finally, and probably most importantly, providing more analysis channels and independent electrodes would provide better perception of overall spectral shape and would more faithfully represent the frequency of the harmonics that are so crucial for music perception.

Another future research direction would be to qualitatively assess the implant's ability to convey musical sounds. It is difficult to ask subjects to rate "pleasantness" or "likeability" with such sounds due to the subjective nature of these kinds of questions. Perhaps an experiment whereby subjects rate "similarity" between the unaltered and processed sounds would reinforce the results obtained in this experiment for the ranking of instruments that are most affected by the implant processing.

One can also experiment with the kinds of signals that are presented to the implant. For speech, it is generally accepted that one can represent intelligible signals with a small set of parameters (for example, in technologies such as LPC analysis [O'Shaughnessy, 1999]). It is still undetermined what might be a "minimal representation" of a music signal. Work done by Ellis [Ellis, 1992] demonstrated how to divide a signal into very small elements that represent the most important part of the sound. These smaller elements could then be re-combined to form a sound (almost) perceptually identical to the original, but synthesized with only a small number of constantly-varying "tracks", consisting of peaks extracted from the input sound. Examples were given whereby high-quality orchestral musical signals were analyzed and re-synthesized with only 20 to 30 sinusoidal tracks at any given slice in time. Ellis' system involved complex computation, with several channels of analysis employing variable sampling rates, followed by extraction of spectral peaks over short windows of time, which might not be conceivable for current implants. Nevertheless, a model such as this one might point researchers in the direction of finding out what minimal amount of information contained in a music signal is necessary for perception. This minimal information could be transmitted more efficiently, instead of the full-bandwidth audio.

As an artistic experiment or application, one might explore the notion of composition for a particular implant. As mentioned earlier, results obtained by Townsend et al. [Townsend et al., 1987] showed that a continuous pitch sensation was obtained by simultaneous stimulation of multiple electrodes. Knowing a particular implanted listener's response to various frequencies, a piece of music could be composed using a mapping of frequencies necessary to elicit perception of the notes produced. This might limit the tonal framework to a small set of notes, but would still be enjoyable and easy to follow.

Similarly, an analysis-by-synthesis paradigm could translate acoustic sounds into meaningful sounds that could be interpreted by the user, knowing the characteristics of the particular implant's processing strategy and the listener's frequency mapping. For example, a piano tone could first be analyzed in time and frequency, as a front-end to the processor. Knowing the target envelope that is required to produce the piano sound, along with the particular harmonic spacing associated with a piano's spectral content, one could stimulate the necessary electrodes in a manner that produces a tone closest to the target sound. This might produce a more faithful representation of the sound, rather than the current system whereby the sound is analyzed and presented "as is", disregarding the fact that frequency content is added or removed. This is one way in which one could work with the current implant and "reverse engineer" a listener's response to synthesize sound appropriately and more faithfully, without drastically altering the implant's design.

The outcome of this research shows that perception of musical instruments is one area that should definitely be addressed in the design of current cochlear implants. As advances in technology occur at such a rapid pace, it is likely that improvements in cochlear implant design will follow suit. Although speech is of most importance for implant recipients to enable fluent communication, music is extremely important for many implant users and should be researched with equal emphasis. The experiment and results described in this thesis, along with the recommendations for future work, should provide a framework for further research in this area.

REFERENCES

(1997). The cochlea. *Encyclopedia Britannica*. http://www.search.eb.com/eb/topic?eu=117508&type=17 (01-08-2002).

(1997). Basilar membrane. *Encyclopedia Britannica*. http://www.search.eb.com/eb/topic?eu=117508&type=17 (01-08-2002).

Bartholomew, Wilmer T. *Acoustics of Music*. New York: Prentice-Hall inc., 1942.

Berger, Kenneth W. Some factors in the recognition of timbre. *J. Acoust. Soc. Am*. 36 (1964): 1888-1891.

Blanchfield, Bonnie B., Feldman, Jacob J., and Dunbar, Jennifer L. (October 1999). The Severly to Profoundly Hearing Impaired Population in the United States: Prevalence and Demographics. *Project HOPE Policy Analysis Brief*. http://www.bionicear.com/support/supp_research_demo2.html (01/08/2002).

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis. *J. Acoust. Soc. Am*. 95 (1994): 2658-2669.

Cohen, A. Further investigation of the effects of intensity upon the pitch of pure tones. *J. Acoust. Soc. Am*. 33 (1961): 1363-1376.

Dorman, Michael F., Loizou, Philipos C., and Rainey, Dawne. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J. Acoust. Soc. Am*. 102 (1997): 2403-2411.

Eagleson, Halson V., and Eagleson, Oran W. Identification of musical instruments when heard directly and over a public-address system. *J. Acoust. Soc. Am*. 19 (1947): 338-342.

Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladejovsky, M.G. and Parkin, J.L., "Place and periodicity pitch by stimulation of multiple scala tympani electrodes in deaf volunteers," *Trans. Am. Soc. Art. Intern. Organs* 24 (1978): 1-5.

Ellis, D.P.E. “A Perceptual representation of audio.” MIT Master’s thesis, 1992.

FDA (Public Health Service). “Re: P000025, The MED-EL COMBI 40+ Cochlear Implant System.” Aug. 20, 2001. *Letter to Mr. A. Thomas Doyle*. 1-2.

Fujita S, and Ito J. Ability of nucleus cochlear implantees to recognize music. *Annals of Otology, Rhinology, and Laryngology*. 108 (1999): 634-640.

Gfeller, Kate, Christ, Aaron, Knutson, John F., Witt, Shelley, Murray, Kathleen T., and Tyler, Richard S. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *J. Am. Acad. Audiol*. 11 (2000): 390-406.

Gfeller, K., and Lansing, C.R. Melodic, rhythmic, and timbral perception of adult cochlear implant users. *Journal of Speech and Hearing Research*. 34 (1991): 916-20.

Gfeller, K., and Lansing, C. Musical perception of cochlear implant users as measured by the *Primary Measures of Music Audiation*: an item analysis. *Journal of Music Therapy*. 29 (1992): 18-39.

Gfeller, Kate, Witt, Shelley, and Mehr, Maureen A. Effects of frequency, instrument family, and cochlear implant type on timbre recognition and appraisal. *Ann. Otol. Rhino. Laryngol*. 111 (2002): 349-356.

Gfeller K, Woodworth G, Robin DA, Witt S, and Knutson JF. Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and Hearing*. 18 (1997): 252-60.

Grey, John M. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am*. 61 (1977): 1270-1277.

Handel, S. Listening: An introduction to the perception of auditory events. Cambridge: MIT Press, 1993.

Helmholtz, H. On the Sensations of Tone as a Physiological Basis for the Theory of Music. Trans. A.J. Ellis. 2nd ed. New York: Dover, 1954.

House, William F. Cochlear Implants. *Ann. Otol*. 85 (1976): 2.

House Ear Institute (2002). The History of Cochlear Implants. http://www.hei.org/hearhealth/hearloss/cihist.htm (29-7-2002).

Kendall, Roger A. The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception*. 4 (1986): 185-214.

Kessel, R.G. and Kardon, R.H. *Tissues and Organs: A Text-Atlas of Scanning Electron Microscopy*, San Francisco: W.H. Freeman and Company, 1979.

Loizou, P.C. Signal-processing techniques for cochlear implants. *IEEE Engineering in Medicine and Biology Magazine*. 18 (1999): 34 –46.

Loizou, Philipos C., Dorman, Michael, and Tu, Shemin. On the number of channels needed to understand speech. *J. Acoust. Soc. Am*. 106 (1999): 2097-2103.

Martin, Keith D., and Kim, Youngmoo E. Musical instrument identification: a pattern-recognition approach. *Presented at the 136th meeting of the Acoustical Society of America* (1998).

Martin, Keith D. Sound-Source Recognition: A Theory and Computational Model. Diss. MIT, June 1999.

McDermott, Hugh J.and McKay, Colette M. Musical pitch perception with electrical stimulation of the cochlea. *J. Acoust. Soc. Am*. 101 (1997): 1622-1631.

Med-El (2002). Understanding cochlear implants. http://www.medel.com/shared/pdf/en/MKT1003E_r2.0_underst-CI.pdf (01-08-2002)

Michelson, R.P., Merzenich, M.M., Schindler R.A., and Schindler, D.N. Present status and future development of the cochlear prosthesis. *Ann. Otol*. 84 (1975): 494.

Moore, Brian C.J. *An Introduction to the Psychology of Hearing*. 3rd ed. San Diego: Academic Press, 1989.

O’Shaughnessy, Douglas. *Speech Communications: Human and Machine*. 2nd ed. New Jersey: John Wiley & Sons, 1999.

Pickles, James O. *An Introduction to the Physiology of Hearing*. 2nd ed. San Diego: Academic Press, 1988.

Pijl S, Schwarz, DWF. Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *J. Acoust. Soc. Am.* 98 (1995): 886-895.

Saldanha, E.L., and Corso, John F. Timbre cues and the identification of musical instruments. *J. Acoust. Soc. Am*. 35 (1964): 2021-2026.

Shannon, Robert V., Zeng, Fan-Gang, Kamath, Vivek, Wygonski, John, and Ekelid, Michael. Speech recognition with primarily temporal cues. *Science, New Series*. 270 (1995): 303-304.

Strong, William, and Clark, Melville Jr. Perturbations of synthetic orchestral wind-instrument tones. *J. Acoust. Soc. Am.* 41 (1966): 277-285.

Tierney, J., Zissman, Mark A., and Eddington, Donald K. Digital signal processing applications in cochlear-implant research. *The Lincoln Laboratory Journal* 7 (1994): 31-62.

Townsend, Brent, Cotter, Neil, Van Compernolle, Dirk, and White, R.L. Pitch perception by cochlear implant subjects. *J.Acoust. Soc. Am*. 82 (1987): 106-115.

Weis, Johannes. A Computer-Based System for the Acoustic Simulation of Auditory Prostheses and Impaired Listeners. Diss. MIT, May 2000.

Wilson, Blake S., Finley, Charles C., Lawson, Dewey T., Wolford, Robert D., Eddington, Donald K., and Rabinowitz, William M. Better speech recognition with cochlear implants. *Nature*. 352 (1991): 236-238.

APPENDIX A: CODE FOR MATLAB SIMULATION

---

rrcisim.m

```
function rrcisim(nch,filename,bpflow,recttype,lpftype,outdir)

% Usage: rrcisim (nch,filename,bpflow,recttype,lpftype,outdir)
%
%        nch -- total number of channels
%        filename - Input filename (you need a directory called "audio_in" for input audio files)
%        bpflow - cutoff of lowest BPF (90 or 350Hz)
%        recttype - rectification ('f' - full, 'h' - half)
%        lpftype - lowpass filter type ('mvav' - 16pt moving average, 'b400' - 400Hz 2nd order butterworth
%        outdir - directory and filename ex. 'real_instruments/piano' (you need a directory called "audio_out")
% -- calls function estfilt.m for filter parameters --
%
% adapted from Loizou (JASA 1999)
%
% Rebecca Reich
% January 2002

global filterA filterB center Srate

%========== open the input file ========== %

% ====== USING WAVREAD (so no need to scale output) ========
[x,Srate,bpsa] = wavread(['audio_in/', filename]);
[row,col] = size(x);
if col==2
              x = sum(x,2);
    disp('summing to mono');
end
% now resample to 12971Hz
if Srate==44100
    x = resample(x,5,17);
    Srate = Srate*5/17;
elseif  Srate == 22050
    x = resample(x,10,17);
    Srate = Srate*10/17;
elseif Srate == 48000
    x = resample(x,459375,1700000);
    Srate = Srate*459375/1700000;
end
disp('downsample to 12971Hz');
```

```
% ===============================================

n_samples=length(x);
nChannels=nch;

% ========== remove any DC bias ==========
x = x - mean(x);

% %--Preemphasize first (LOIZOU method) ---------
% bp = exp(-1200*2*pi/Srate);
% ap = exp(-3000*2*pi/Srate);
% x = filter([1 -bp],[1 -ap],x); %using freqz: cuts off high freqs?


% ========== get bandpass filter coefficients ==========
if isempty(filterA)
   estfilt(nChannels,bpflow);
   fprintf('\n Getting filter coefficients for %d-channel processor..\n',nch);
end

% ========== calculate lowpass filter coefficients  ==========
if lpftype == 'mvav'
                mvavgord = 16; % sample length of moving average filter (LPF smoothing filter)
                blow = 1/(mvavgord) * ones(1,mvavgord);
                alow = 1;
elseif lpftype == 'b400'
   [blow,alow]=butter(2,400/(Srate/2)); % b400 has 2nd order filter for signals done at 44.1/22.05kHz
   %[blow,alow] = butter(6,400/(Srate/2));
else
   [blow,alow]=butter(2,40/(Srate/2)); % an attempt at a lower cutoff frequency
end
% ========== filter input with BPFs, lowpass filter the rectified output ==========
ylpf=zeros(nChannels,n_samples);
ybpf=zeros(nChannels,n_samples);

for i=1:nChannels
   ybpf(i,:)=filter(filterB(i,:),filterA(i,:),x)'; % bandpass
   if recttype == 'f'
      yrect(i,:) = abs(ybpf(i,:)); % full-wave rectify
      ylpf(i,:)=filter(blow,alow,yrect(i,:)); % lpf
   else % half-wave
      yrect(i,:) = ybpf(i,:);
      yrect(i,yrect(i,:)<0) = 0;    % set neg. values to zero
      ylpf(i,:) = filter(blow,alow,yrect(i,:)); % lpf
   end
end
%ysum = sum(ylpf,1); %for playing purposes
disp('done filtering');


% HIGHPASS (LOIZOU method) ----------------
% [bhi,ahi] = butter(4,20/(Srate/2));
% for i = 1:nChannels
%     ylpf(i,:) = filter(bhi,ahi,ylpf(i,:));
% end
% -----------------------

% ========== outputs ==========
freq=center/Srate; %normalized center frequencies
for i=1:nChannels % ------ generate sinewaves ---------
      ycos(i,:)=cos(2*pi*freq(i)*[0:n_samples-1]);
      yout_off(i,:)=ylpf(i,:).*ycos(i,:);  %modulate sinewaves with envelopes
      yout(i,:) = yout_off(i,:);
end
youtsum = sum(yout,1);
youtsum = youtsum/(max(abs(youtsum)));

if max(abs(youtsum))>1, fprintf('Warning! Overflow in file %s\n',filename); end;
%--- save output to a file ----
wavname = ['_ch',num2str(nch),'_BPF',num2str(bpflow),'_RECT',recttype,'_LPF',num2str(lpftype),'.wav'];
name = ['audio_out/',outdir,wavname]
wavwrite(youtsum,Srate,bpsa,['audio_out/',outdir,wavname]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% TIME AND FREQUENCY VECTORS FOR PLOTTING
% tvec = 0:1/Srate:(n_samples-1)/Srate;
% fvec = [0:n_samples-1]./n_samples*Srate;
```

estfilt.m

```
function estfilt(nChannels,bpfcutoff)

% Function: estfilt.m
%
% estimate the parameters of the bandpass filters--
%
% Advanced Bionics, Clarion, Platinum Sound Processor Filter Table (2000) CIS
%
% input: nChannels - number of channels
%
% adapted from Loizou's estfilt.m
%
% Rebecca Reich
% January 2002

global filterA filterB center Srate

SAVE=0;  % if 1, save center frequencies and bandwidths in a file
FS=Srate/2;

nOrd=6;
center=zeros(1,nChannels);

% ========== frequency boundaries for bandpass filters, depending on # of channels ========== %
if nChannels == 16
   % ============ 16-channels, logarithmically-spaced ======================
            UpperFreq=6800; LowFreq=bpfcutoff;
            range=log10(UpperFreq/LowFreq);
            interval=range/nChannels;
            center=zeros(1,nChannels);
            for i=1:nChannels  % ----- Figure out the center frequencies for all channels
            upper1(i)=LowFreq*10^(interval*i);
                        lower1(i)=LowFreq*10^(interval*(i-1));
            end
else
            switch nChannels
            case 1
                        upper1 = 6800;
                        lower1 = bpfcutoff;
            case 2
     upper1 = [1387 6800];
     lower1 = [bpfcutoff 1387];
            case 3
     upper1 = [877 2196 6800];
     lower1 = [bpfcutoff 877 2196];
            case 4
     upper1 = [697 1387 2762 6800];
     lower1 = [bpfcutoff 697 1387 2762];
            case 5
     upper1 = [607 1053 1827 3170 6800];
     lower1 = [bpfcutoff 607 1053 1827 3170];
            case 6
     upper1 = [554 877 1387 2196 3475 6800];
     lower1 = [bpfcutoff 554 877 1387 2196 3475];
            case 7
     upper1 = [519 769 1140 1689 2504 3711 6800];
     lower1 = [bpfcutoff 519 769 1140 1689 2504 3711];
            case 8
     upper1 = [494 697 983 1387 1958 2762 3898 6800];
     lower1 = [bpfcutoff 494 697 983 1387 1958 2762 3898];
            end
end
for i = 1:nChannels % take GEOMETRIC MEAN (not simple average)
   center(i)=sqrt(upper1(i)*lower1(i));
end

% ========== for SAVE = 1 print filter values to a file ========== %
if SAVE==1
   fps=fopen('filters.txt','a+');
   fprintf(fps,'%d channels:\n',nChannels);
   for i=1:nChannels
      fprintf(fps,'%d ',round(upper1(i)-lower1(i))); %bandwidths
   end;
```

```
    fprintf(fps,'\n');
    for i=1:nChannels
        fprintf(fps,'%d ',round(center(i))); %center frequencies
    end;
    fprintf(fps,'\n=======\n');
    fclose(fps);
end

% ========== check for aliasing ========== %
if FS<upper1(nChannels) % need sRate >= 2 * max freq
    useHigh=1;
else
    useHigh=0;
end

% ========== design the filters ========== %
filterA=zeros(nChannels,nOrd+1);
filterB=zeros(nChannels,nOrd+1);
for i=1:nChannels
    W1=[lower1(i)/FS, upper1(i)/FS];
    if i==nChannels
        if useHigh==0
            [b,a]=butter(0.5*nOrd,W1);
        else
            [b,a]=butter(nOrd,W1(1),'high');
        end
    else
        [b,a]=butter(0.5*nOrd,W1);
    end
    filterB(i,1:nOrd+1)=b;
    filterA(i,1:nOrd+1)=a;
end
```

APPENDIX B: MUSICAL BACKGROUND SURVEY

1. Please list any instruments you have played or currently play, including the number of years you were/are actively playing the instrument on a regular basis.

Instrument(s)	# years played	last year when actively playing

2. Please indicate the number of years you had private lessons on the instrument(s) and the last time you had regular lessons:

Instrument(s)	# years of lessons	last year when taking regular lessons

3. Please indicate any ensembles with which you have performed, including the nature of the ensemble, length of time you were involved, and the year you last performed with the ensemble.

Ensemble	# years involved	last year performed

4. Please describe any theory, ear-training, music appreciation, composition or related courses you may have taken, including a short description of the content and the year(s) when taken.

Course	Year(s) taken

Please classify yourself into one of the following categories (circle the most appropriate number):

1. No formal training, little knowledge about music, and little experience in listening to music

2. No formal training or knowledge about music, but informal listening experience

3. Self-taught musician who participates in musical activities

4. Some musical training, basic knowledge of musical terms, and participation in music classes or ensembles in elementary or high school

5. Several years of musical training, knowledge about music, and involvement in music groups

AGE (please circle one):

<16 years

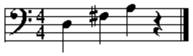
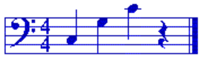
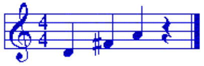
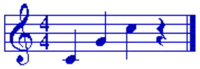
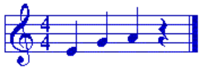
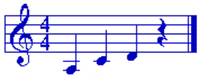
17-30 years

31-50 years

51+ years

Do you have any known hearing problems? Yes: _____   No: _____

APPENDIX C: SYNTHESIZED INSTRUMENT STIMULI

This table shows the musical sequences composed for synthesized instruments. The shaded boxes indicate that the particular sequence was played by the instrument indicated on the left-hand side. Note that transcriptions of the real instruments stimuli were not available; recorded excerpts were mostly taken from Martin's database.

	 sequence 1	 sequence 2	 sequence 3	 sequence 4	 sequence f	 sequence all
Bassoon	▨	▨				▨
Oboe			▨	▨		▨
Clarinet			▨	▨		▨
Saxophone			▨	▨		▨
Flute			▨	▨	▨	
Trumpet			▨	▨		▨
Trombone	▨	▨				▨
Violin			▨	▨		▨
Cello	▨	▨				▨
Guitar	▨	▨				▨
Piano			▨	▨		▨