

Published in final edited form as:

Hear Res. 2011 October ; 280(1-2): 192–200. doi:10.1016/j.heares.2011.05.017.

Impaired perception of temporal fine structure and musical timbre in cochlear implant users

Joseph Heng^a, Gabriela Cantarero^a, Mounya Elhilali^b, and Charles J. Limb^{a,c,*}

^aDepartment of Otolaryngology-Head and Neck Surgery, Johns Hopkins Hospital, Baltimore, MD, USA

^bDepartment of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

^cPeabody Conservatory of Music, Johns Hopkins University, Baltimore, MD, USA

Abstract

Cochlear implant (CI) users demonstrate severe limitations in perceiving musical timbre, a psychoacoustic feature of sound responsible for ‘tone color’ and one’s ability to identify a musical instrument. The reasons for this limitation remain poorly understood. In this study, we sought to examine the relative contributions of temporal envelope and fine structure for timbre judgments, in light of the fact that speech processing strategies employed by CI systems typically employ envelope extraction algorithms. We synthesized “instrumental chimeras” that systematically combined variable amounts of envelope and fine structure in 25% increments from two different source instruments with either sustained or percussive envelopes. CI users and normal hearing (NH) subjects were presented with 150 chimeras and asked to determine which instrument the chimera more closely resembled in a single-interval two-alternative forced choice task. By combining instruments with similar and dissimilar envelopes, we controlled the valence of envelope for timbre identification and compensated for envelope reconstruction from fine structure information. Our results show that NH subjects utilize envelope and fine structure interchangeably, whereas CI subjects demonstrate overwhelming reliance on temporal envelope. When chimeras were created from dissimilar envelope instrument pairs, NH subjects utilized a combination of envelope ($p = 0.008$) and fine structure information ($p = 0.009$) to make timbre judgments. In contrast, CI users utilized envelope information almost exclusively to make timbre judgments ($p < 0.001$) and ignored fine structure information ($p = 0.908$). Interestingly, when the value of envelope as a cue was reduced, both NH subjects and CI users utilized fine structure information to make timbre judgments ($p < 0.001$), although the effect was quite weak in CI users. Our findings confirm that impairments in fine structure processing underlie poor perception of musical timbre in CI users.

1. Introduction

Individuals with cochlear implants (CI) frequently struggle with the perception of musical stimuli. In addition to well-described impairments in pitch processing (Moore and Carlyon, 2005), CI users display severely limited abilities in the assessment of musical timbre, which is the core focus of this study. Timbre, or ‘tone color’, is defined as the set of attributes that

© 2011 Elsevier B.V. All rights reserved.

*Corresponding author. 720 Rutland Ave, Ross 826, Baltimore, MD 21205, USA. Tel.: +1 410 502 4269; fax: +1 410 955 0035. climb@jhmi.edu (C.J. Limb).

Competing interests: Charles Limb is a consultant for Advanced Bionics Corporation, a manufacturer of cochlear implants, and receives support for unrelated work.

allows a listener to differentiate between musical instruments playing at the same pitch, amplitude and duration (Ansi, 1973). It is also essential for both the cognitive and aesthetic aspects of music, which often contains multiple streams of information with widely varying spectral and temporal characteristics (Caclin et al., 2006) that are distinguished primarily by instrumental timbre. While several studies have described poor performance of CI users during timbre identification tasks (Gfeller et al., 2002a,b; McDermott, 2004; Nimmons et al., 2008), the reasons for this poor performance remain unclear.

For over a century, the property of timbre was associated with the distribution of spectral energies within a sound (Von Helmholtz and Ellis, 1895). Multidimensional scaling models have been applied to determine the perceptual components of timbre (Grey, 1977; Krumhansl, 1989; Marozeau et al., 2003; McAdams et al., 1995; Samson et al., 1997), the most important of which are temporal envelope modulation and spectral distribution of the harmonic frequencies of sound (fine structure). By the Hilbert transform, the envelope can be mathematically defined as the magnitude of the analytic signal, while fine structure can be defined as the cosine of the phase of the analytic signal.

Our goal in this study was to examine how normal hearing listeners and CI users utilize fine structure and envelope information during timbre discrimination. Due to the use of implant-based speech processing strategies that emphasize envelope detection and discard fine structure information, it has been suggested that CI users rely solely upon envelope cues during timbre judgments (Kong et al., 2004), while individuals with normal hearing are thought to utilize both envelope and fine structure (Gunawan and Sen, 2008; Kong et al., 2004; Smith et al., 2002). In addition, implant-based speech processing strategies have forced CI users to utilize a limited number of frequency bands in auditory perception.

In this study, we created “instrumental chimeras” that were synthesized from multiple pairs of instruments and represent musical hybrids in terms of timbre. These chimeras contained variable proportions of envelope and fine structure from each source instrument used to generate the chimera, allowing us to assess the relative contributions of envelope and fine structure to timbre identification. Earlier work using auditory chimeras has demonstrated the critical importance of fine structure to melody identification, in 8 frequency bands or less (Smith et al., 2002). Furthermore, Xu and Pfingst have underscored the importance of fine structure cues for lexical tone perception when 4–16 frequency bands were used (Xu and Pfingst, 2003).

In this study, we hypothesized that CI users would not utilize fine structure information during timbre judgments, in comparison to normal hearing subjects, who were predicted to rely on both envelope and fine structure cues.

2. Methods

2.1. Stimuli

Original instrument samples that served as source files for chimera synthesis were recorded using the Miroslav Philharmonik Suite (IK Multimedia), and Ivory Grand Pianos (Synthogy) on the Apple Logic Pro 7.0 platform. Instrumental chimeras were created using a custom MATLAB-based chimera synthesis program (after Smith et al., 2002) [MATLAB R2007a by Mathworks]. Four instruments playing an identical eight-note novel melody were used to generate these chimeras. Of these four instruments, two had percussive envelopes (piano, guitar), while the other two had sustained (flute, trumpet) envelopes. These four instruments were chosen to represent these percussive and sustained classifications as opposed to other instruments due to their common usage in music, and also because the percussive envelopes of the piano and guitar did not have abrupt temporal decay, which would have complicated

chimera synthesis (signal durations were equal for all source instruments). Similar envelope chimeras were created from instrument pairs with similar envelopes (percussive/percussive or sustained/sustained), while dissimilar envelope chimeras were created from instrument pairs with dissimilar envelopes (percussive/sustained).

The program constructed auditory chimeras by using the Hilbert transform to extract the Hilbert envelope and fine structure from the analytic signals of two selected instrument samples, and then recombining them in different ratios to construct an instrumental chimera (Fig. 1). A total of 100 chimeras were created and presented from instrument source pairs with dissimilar envelopes. These chimeras were created in order to permit the utilization of both envelope and fine structure cues. In addition, a total of 50 chimeras were created and presented from instrument source pairs with similar envelopes. Unlike the chimeras created from dissimilar envelopes, these chimeras were created in order to limit the extent to which envelope cues could be utilized. Finally, for further analysis, we looked at 18 chimeras that contained contradictory information (i.e. dominant envelope from instrument A but dominant fine structure from instrument B) that were considered “ambiguous”. These chimeras are a subset of the 150 chimeras generated for subject testing. These stimuli contained chimeras that were composed of 50:50 envelope and 50:50 fine structure representation ratios, 75:25 envelope and 25:75 fine structure representation ratios (and vice versa), and 100:0 envelope and 0:100 fine structure ratios (and vice versa) of the two source instruments.

To create an instrumental chimera, we used the analytic signal $a(t) = a_o(t) + ia_h(t)$, where $a_o(t)$ is the output of the source file used in chimera synthesis, $a_h(t)$ is the Hilbert transform of $a_o(t)$, and $i = \sqrt{-1}$. The Hilbert envelope is the magnitude of the analytic signal,

$m(t) = \sqrt{[a_o^2(t) + a_h^2(t)]}$. The fine structure is the cosine of the phase of the analytic signal, $\phi(t) = \arctan(a_h(t)/a_o(t))$. The chimera can be constructed as $c(t) = [x*m_1(t) + y*m_2(t)]*[x*\cos \phi_1(t) + y*\cos \phi_2(t)]$, where $x + y = 1$, and x and y are the desired percentage distributions of the Hilbert envelope and fine structure of the two selected instrument samples. Since cochlear implant speech processors work via spectral filtering of various frequency bands, the chimeras created here were not divided into an arbitrary number of filter bands prior to application of the Hilbert transform. Rather, the entire signal was treated as one broadband to which the Hilbert transform was applied in its entirety.

As a simple verification check to ensure that the chimerizer was functioning properly, we were able to faithfully reconstruct our original source files through the chimerizer (after envelope and fine structure extraction and then recombination). This was simply the extraction of a sample's Hilbert envelope and its fine structure and recombining them to reform the original sample. All auditory stimuli were normalized by root-mean-square power.

In addition, a gammachirp filter bank simulation identical to Gilbert and Lorenzi (2006) was built to examine the extent of envelope recovery of the chimeras at the output of six gammachirp auditory filters. These 6 gammachirp filters are a representation of the limited number of filters a CI recipient can employ in envelope recovery. The signal is band pass filtered between 80 and 8020 Hz using Butterworth filters (for the case of 1 band, which is what we used here). The fine structure is then extracted using the Hilbert transform (this is known as the HFS signal). The original and HFS signals are then passed through 6 gammachirp filters. A lowpass Butterworth (forward and backward) filter is applied to the envelope. The mean correlation coefficients between the original envelopes and the recovered envelopes of the chimeras are then computed at the output of six gammachirp auditory filters using MATLAB's *corrcoef* function.

2.2. Subjects and test procedure

The target test population consisted of NH listeners ($n = 14$; mean age 26.1 ± 5.2 years) and CI users ($n = 12$; mean age 53.3 ± 13.5 years) (Table 1). All CI users were post-lingually deafened adults. All subjects completed a musical experience questionnaire to ascertain the extent of their musical training; no subjects had formal musical training beyond the amateur level. All experiments were performed at the Sound and Music Perception Laboratory of Johns Hopkins Hospital, and carried out after the review and approval of the Johns Hopkins Hospital Institutional Review Board. Informed consent was obtained for all subjects. A brief training session took place before the actual test to familiarize the subjects with the test procedure and original instrument samples. All stimuli were played free-field through a calibrated loudspeaker (Sony SS-MB150H) in a sound booth at a presentation level of 80 dB HL through an OB822 clinical audiometer (Madsen Electronics). Any ears with residual hearing were occluded with an ear plug, and no hearing aids were worn in non-implanted ears. Bilateral implantees ($n = 2$) were tested using only their first implant. All CI subjects used their everyday speech processors during the experiment.

All chimeras generated from one pair of instruments were presented serially, but in randomized order. Subjects were informed of the two source instruments (e.g. flute or piano, guitar or trumpet, etc.) from which they were asked to choose the source instrument that they felt was most similar to the presented stimulus, in a two-interval forced choice alternative task. Each stimulus lasted 5 s, followed by a response period of 5 s. Subjects were instructed to guess if unsure. All subjects were presented with all 150 generated chimeras. The entire test paradigm lasted for 45 min. There were no null responses recorded.

3. Results

3.1. Chimeras created from dissimilar envelope instruments

A total of 100 chimeras were created and presented from instrument source pairs with dissimilar envelopes. These chimeras were created in order to permit the utilization of both envelope and fine structure cues. Fig. 2 shows timbre judgments made by both subject groups in response to these chimeras, with the x -axis representing the ratio of fine structure for instruments A and B from 100:0 to 0:100 in 25% increments (where A and B represent any source instrument pairs with dissimilar envelopes, e.g. piano/flute or trumpet/guitar) and the y -axis representing the percentage of times (from 0 to 100%) that the subject selected instrument B as most closely resembling the presented chimera. The different colored lines represent chimeras of different envelope ratios between source instruments A and B, from 100:0 to 0:100 also in 25% increments. The blue line represents chimeras with exactly 50% envelope representation from each source instrument and no theoretical bias for either instrument A or B. As shown in the graph for CI subjects, responses to the 50:50 envelope chimera (blue line) were clustered around chance regardless of the fine structure components of the chimera, whereas normal hearing subjects display results around chance for the same chimeras only when the ratio of fine structure was also exactly 50:50 in representation. Similarly, CI subjects consistently identified chimeras with 100:0 envelope representation as representing instrument A, even when fine structure representation was completely reversed at 0:100 for source instruments A and B. Overall, CI users showed a much greater reliance on envelope cues than fine structure information in making timbre judgments for similar envelope chimeras. Statistical analysis using a two-way repeated-factor ANOVA analysis revealed a significant effect for envelope ($p < 0.001$) but not for fine structure ($p = 0.908$) (Fig. 2). In contrast, NH subjects used both envelope and fine structure information in timbre identification, with statistical analysis by two-way repeated-factor ANOVA showing a significant effect for both envelope ($p = 0.008$) and fine structure ($p = 0.009$). No significant interactions were found.

In previous studies of timbre perception in CI users, CI users identified percussive instruments more readily than wind or string instruments due to the greater distinctiveness of the temporal envelopes in the percussive instruments (McDermott and Looi (2004); Nimmons et al., 2008). To look for a similar result in this study, we examined timbre judgments in three scenarios: chimeras that were composed of 100% envelope of a percussive instrument and 100% fine structure of a sustained instrument, chimeras that had envelope and fine structure weighted equally between a percussive and a sustained instrument, and chimeras that were composed of 100% envelope of a sustained instrument and 100% fine structure of a percussive instrument. In the second scenario where chimeras that had envelope and fine structure weighted equally, timbre judgments were even at 50%. In the first scenario where envelope information was weighted entirely toward the percussive instrument, CI subjects chose the percussive instrument 93.8% of the time. In the third scenario where envelope information was weighted entirely toward the sustained instrument, they chose the sustained instrument 79.2% of the time. However, the difference between the timbre judgments in the first and third scenario was not significant (t -test, $p > 0.01$).

3.2. Chimeras created from similar envelope instruments

A total of 50 chimeras were created and presented from instrument source pairs with similar envelopes. Unlike the chimeras created from dissimilar envelopes, these chimeras were created in order to limit the extent to which envelope cues could be utilized. As above, Fig. 3 shows timbre judgments (shown here as percentage of times that source instrument B was selected) for both subject groups as a function of fine structure representation ratio (x -axis) and envelope representation ratio (different colored lines). This figure shows that both groups judged the chimeras similarly independent of envelope ratio, and neither NH controls nor CI users displayed a statistically significant utilization of envelope information for timbre judgments of similar envelope chimeras. Interestingly, CI users demonstrated a statistically significant effect of fine structure representation ratio on timbre judgments (two-way ANOVA repeated-factor, $p < 0.001$), similar to NH controls (two-way ANOVA, repeated-factor, $p < 0.001$). No significant interactions were found.

A direct graphical comparison between the responses of NH controls and CI users showed that fine structure information had a smaller influence on the timbre judgments of CI users than NH subjects (Fig. 3).

3.3. Ambiguous chimeras

A total of 18 chimeras were generated that contained contradictory information (i.e. dominant envelope from instrument A but dominant fine structure from instrument B) that were considered “ambiguous”. These chimeras are a subset of the 150 chimeras generated for subject testing. These chimeras were used to examine how CI subjects would utilize envelope and fine structure cues in ambiguous situations. These stimuli contained chimeras that were composed of 50:50 envelope and 50:50 fine structure representation ratios, 75:25 envelope and 25:75 fine structure representation ratios (and vice versa), and 100:0 envelope and 0:100 fine structure ratios (and vice versa) of the two source instruments. An analysis of timbre judgments for such ambiguous chimeras generated from source instruments with dissimilar envelopes revealed that CI users relied more on envelope cues than normal hearing subjects (Fig. 4, dark bars), even when fine structure information directly contradicted this judgment. CI subjects selected source instruments consistent with the dominant envelope in 86% of cases for 100:0 envelope and 0:100 fine structure ratios, and 72% of cases for 75:25 envelope and 25:75 fine structure ratios. When presented with ambiguous chimeras synthesized from source instruments with similar envelopes, there was no significant difference among controls or CI users’ judgments (t -test, $p > 0.01$; Fig. 4, light

bars). In a perfectly ambiguous situation in which envelope and fine structure information were equally weighted between two source instruments (50:50 for both envelope and fine structure), the timbre judgments of both normal hearing adults and CI users were at chance level (Fig. 4, right). These results suggest that unlike normal hearing controls, CI subjects strongly favor envelope information when available, even when this information is directly contradicted by the fine structure information provided.

3.4. Responses according to source instruments

We analyzed subject responses according to specific source instruments, to evaluate whether or not a subset of instruments were “driving” the results statistically. As shown in Fig. 5, there are minor differences in subject responses for each specific instrument. However, overall patterns for each instrument remain similar to those described above, with a statistically small yet significant influence of fine structure information on timbre judgments for CI users (two-way ANOVA, repeated-factor, p -values <0.01). However, for each indicated instrument, normal hearing subjects displayed much more significant utilization of fine structure information than CI users (two-way ANOVA, repeated-factor, p -values <0.001).

3.5. Reconstruction of envelope cues

To examine the possibility of recovered envelope cues affecting subjects’ timbre judgments, we used a gammachirp filter bank simulation identical to Gilbert and Lorenzi (2006) to compute the mean correlation coefficients between the original envelopes and the recovered envelopes of the chimeras at the output of six gammachirp auditory filters (21). Correlation coefficients were below 0.5, indicating no significant resemblance between the original envelopes and those recovered at the output of auditory filters (Fig. 6).

4. Discussion

In this study, we used instrumental chimera synthesis to examine the perception of music timbre. We utilized a test method that examined the basis of subject responses rather than a performance scale (e.g. percent correct on a timbre identification task) in order to conclude that limitations in fine structure processing contribute to, or at least, are in part responsible for poor timbre perception in CI users.

In addition, we found that CI users, unlike normal hearing controls, displayed an overwhelming reliance on envelope cues for timbre judgments. Even in cases when the presented chimera contained none of the fine structure of a given source instrument, CI subjects consistently selected that instrument if the envelope was dominant. In these cases, no significant utilization of fine structure cues was displayed. This supports findings that CI users predominantly rely on envelope cues during timbre evaluation. In studies by McDermott and Looi (2004) and Nimmons et al. (2008), CI users identified percussive instruments more readily than wind or string instruments due to the greater distinctiveness of the temporal envelopes in the percussive instruments (McDermott and Looi (2004); Nimmons et al., 2008). In these studies, CI users were able to rely on a much greater amount of envelope information (in these cases, the original percussive envelopes of the instruments). However, in the ambiguous situation where envelope and fine structure information were equally weighted between a percussive instrument and a sustained instrument, the percussive envelope information was reduced by 50%, possibly greatly reducing CI users’ ability to rely on temporal cues to make timbre judgments. This heavy reliance on envelope information may in fact be responsible for much of the difficulties in timbre perception faced by CI users (Gfeller et al., 2000). By comparison, normal hearing subjects, when presented with confounding envelope information (that is, with envelope

information that was indistinguishable) relied upon the fine structure information that was available for timbre judgments, and also interchangeably used envelope and fine structure cues as available. This supports the findings of others that normal listeners use fine structure during instrument discrimination (Smith et al., 2002).

In light of the preference for envelope-based timbre judgments, an interesting result from this study shows that CI users appeared to be able to utilize a limited amount of fine structure information for timbre discrimination, as revealed by an analysis of similar envelope chimeras and by modeling of recovered envelope cues at the output of a gammachirp filter bank. This unexpected finding contradicts our initial hypothesis that CI users would be unable to rely on fine structure information to make timbre judgments. Given the fact that fine structure information is putatively removed in implant processing strategies, this finding is surprising. In traditional cochlear implant processing strategies, which are optimized for speech, contiguous band pass filters extract envelope cues from an incoming signal, which are then mapped using electrical pulses to an intracochlear electrode (Wilson, 2004). This sole transmission of envelope cues has been found to be adequate for providing high levels of speech perception in quiet (Friesen et al., 2001). It should be mentioned that none of our subjects used special speech processing strategies designed to preserve fine structure information.

There are several reasons why fine structure information transmission is difficult for a cochlear implant, even if envelope detection strategies are not employed. First, the high stimulation frequencies required for the transmission of fine structure result in the degradation of phase-locking of the auditory nerve (Joris and Yin, 1992). Second, the relative phase of response along the basilar membrane has been observed to shift over time (Reiss et al., 2007; Shamma and Klein, 2000), leading to a mismatch between place of transmission and site of encoding for fine structure along the basilar membrane (Huss and Moore, 2005). Our results suggest that fine structure processing in CI users may exist to a limited degree or in some impoverished form. This finding is supported by a recent study of CI users using envelope-modulated speech processing in which it was found that cochlear implantees showed a limited ability to perceive fine structure cues (Ruffin et al., 2007). Given that CI strategies use envelope extraction, thereby largely discarding fine structure information, CI users may be utilizing residual capacity to resolve changes in fine spectral details. One way in which this might occur is the utilization of broadband temporal fine structure cues through a typical narrow-band envelope processing strategy.

The case for the role of fine structure in sound perception is not limited to musical timbre. Sheft et al. presented evidence that fine structure conveys important phonetic speech information that is independent of any envelope reconstruction that might occur due to auditory filtering (Sheft et al., 2008). In addition, Xu and Pfingst demonstrated the importance of fine structure in tonal language perception (Xu and Pfingst, 2003). Further evidence suggesting that temporal fine structure cues carry relevant information for sound identification and discrimination has been suggested by a number of investigators in a wide array of approaches (Hong and Rubinstein, 2003a, 2006; Hong et al., 2003b; Jolly et al., 1996; Laneau et al., 2006; Litvak et al., 2003a, 2003b; Nogueira, 2005; Oxenham et al., 2004; Shepherd and Javel, 1999).

4.1. Other considerations

Timbre is a complex psychoacoustic feature of music that remains difficult to define quantitatively, and therefore, difficult to measure. Although we approached timbre from the components of temporal envelope and fine structure, it should be mentioned that no clear consensus exists as to the exact definitions of these acoustic properties and precisely how they relate to one another. We selected a mathematical approach to envelope and fine

structure based on the Hilbert transform, which allowed us to derive quantifications of each component based on the analytic signal that could be subsequently recombined in novel ratios. Other studies have attempted to present a temporally-based definition of envelope and fine structure, classifying them as the set of frequencies between 2 and 50 Hz and above 500 Hz, respectively (Plomp, 1983; Rosen, 1992). This controversy regarding the nature of envelope and fine structure is further complicated by findings from Ghitza et al., who demonstrated that normal hearing individuals presented with envelope-filtered auditory stimuli were able to reconstruct spectral cues in the auditory system (Ghitza, 2001). Further studies have shown an innate ability of the auditory system to recover the narrow-band envelope structure from broadband fine structure information (Licklider and Pollack, 1948; Zeng et al., 2004). Recently, it was demonstrated that normal listeners were able to recover envelope cues from speech fine structure (Gilbert and Lorenzi, 2006). Taken together, these studies suggest that it may not be possible to truly isolate envelope or fine structure and that further studies are needed.

The issue of disentangling envelope from fine structure components is a contentious one. Its complication stems from two main issues. The first issue has to do with signal processing principles, whereby envelope and fine structure (carrier signal) of band-limited signals are mathematically dependent: It has been analytically shown that fine structure can be recovered from envelope (Logan, 1977; Papoulis, 1983; Voelcker, 1966). This point is mostly pertinent given the narrow-band filtering taking place at the auditory periphery. In this regard, Gilbert and Lorenzi argued that cochlear filtering effectively maps instantaneous frequency modulation (FM) at the output of each sub-band into amplitude modulations (AM) corresponding to envelope fluctuation (Gilbert and Lorenzi, 2006). Given the interdependence between the envelope and fine structure components of band-limited signals, investigating the individual role of one or the other in perception has become challenging (Sheft et al., 2008; Zeng et al., 2004). The second issue is due to the use of Hilbert envelopes in segregating envelope and fine structure. The magnitude of the analytic signal of a band-limited signal (aka. envelope component) has long been known to expand beyond the nominal bandwidth of the original signal (Dugundji, 1958). Shimmel and Atlas have in fact argued that the Hilbert envelope approach does not satisfy the 'bandwidth invariance' property; whereby derived envelope and fine structure signals tend to have larger bandwidths than the original sub-band signal (Schimmel and Atlas, 2005).

In a well-known study, Smith et al. (2002) manipulated the envelope and fine structure components of speech and music to form chimeras that revealed dichotomies in auditory perception in normal subjects, with envelope being found to be critical for speech perception and fine structure for music perception (Smith et al., 2002). Subsequent studies utilized a similar approach to measure auditory perception in other areas, such as lexical tone perception (Liu and Zeng, 2006; Xu and Pfungst, 2003). We used a similar approach here within the musical domain, by creating instrumental chimeras that would allow us to examine how cochlear implant subjects perceive musical timbre. Where Smith et al. examined auditory dichotomies on the basis of frequency bands and intact envelope and fine structure information, we examined them with different combinations of envelope and fine structure information from two given instruments (Smith et al., 2002).

There are intrinsic limitations on how to present identical auditory stimuli both normal hearing listeners and CI users in an explicitly comparable fashion. In previous studies of auditory chimeras, frequency bands were used to simulate auditory filters in normal hearing. In these studies, a Hilbert transform of each frequency band was applied, with chimerization of each individual frequency band and recombination of all bands into a composite signal. In the case of CI users, speech processing strategies employed on a daily basis typically rely on this method of filter bank analysis and processing (Wilson, 2004), with subsequent

distribution of the processed output to a particular electrode of choice thought to correspond topically in location to the desired frequency band. To avoid issues that would result from mismatched filter banks being successively applied (first by narrow-band Hilbert transform with chimerization and second by speech processor extraction), we decided here to consider the most realistic situation during listening, and used identical free-field auditory stimuli presented to both groups created from wideband Hilbert transformation with chimerization. There are several potential issues with this approach, which we address here. Several studies of speech show that envelope recovery occurs when broadband analysis filters are used prior to stimulus presentation for normal hearing subjects (Gilbert and Lorenzi 2006; Zeng et al., 2004). To reduce the potential implications of envelope reconstruction in the auditory system, we synthesized chimeras from both similar and dissimilar envelope source instruments. As a result, we were able to examine timbre judgments that might take place *even in the case of envelope reconstruction*, by reducing the value of envelope as a cue in similar envelope chimeras. Likewise, dissimilar envelope chimeras allowed us to increase the value of envelope as a cue, and present it in competing fashion against a variable range of fine structure information. In our study, both CI and NH subject responses were distributed at 50% for chimeras synthesized with envelopes and fine structure distributed equally between two source instruments. This distribution of responses around the chance level for this particular type of chimera constitutes an important verification point that argues against the notion that reconstructed envelope cues were being utilized for timbre judgments.

Since we used wideband stimuli to examine CI perception of chimeras, our approach may limit a direct comparison of the findings we present here and previous studies performed using auditory chimeras. Our results should therefore be taken with caution. Further experiments utilizing narrow-band stimuli and examining how narrow-band stimuli are re-filtered in individual CI programming maps could be carried out to verify our findings. Although we minimized potential envelope reconstruction from fine structure by presenting confounding envelope cues in similar envelope chimeras, there is the possibility that CI users may utilize residual capacity at detecting changes in fine spectral details. One other consideration is that all of the CI participants in this study, except for CI2, were high performing recipients, having HINT-Q scores of above 90%. There is indication in existing research that higher-performing recipients have better psychoacoustic perception results in some tests (e.g. there was a strong correlation between word-recognition ability and melody identification ability) (Gfeller et al., 2002a,b). However, upon further analysis, there were no correlations between the HINT-Q or HINT-N scores and the measures of FS perception. Obviously, the minimization or reduction of envelope cues is not the same as elimination, and it is worth considering here that it might not be truly possible to separate envelope information from fine structure information in either theory or practice.

Another question raised in this study was the reliability of our results. If CI participants could not reliably discriminate between the 4 original instruments, it would bring into question their responses when asked to select the instrument most representative of the auditory chimeras. Of the twelve CI users that participated, one CI recipient was unable to discriminate between the piano and guitar, and another CI recipient was unable to discriminate between the flute and trumpet. Given that these instruments had similar envelopes, it was expected that a few CI subjects would be unable to discriminate between them. Nevertheless, the clear majority of our CI recipients were able to assess the instrument correctly.

5. Conclusion

Our results demonstrate that fine structure processing exists in CI users in some impoverished form, even though implant-based speech processing strategies essentially remove these cues. Nevertheless, the possibility remains that the apparent fine structure detection we observed really reflects a form of envelope recovery for chimeras in which source envelopes were dissimilar (similar source envelope chimeras obviated the effects of any envelope reconstruction since they cannot be distinguished from one another). The ability to utilize this impoverished form of fine structure processing through training may potentially lead to better timbre perception in CI users. In addition, fine structure processing should be improved in current implant-based processing strategies to improve timbre perception in CI users. Ultimately, the findings here suggest critical importance of temporal fine structure information for proper musical timbre perception and highlight the significant limitations of CI users to perceive such information.

References

- Ansi, PT. Psychoacoustical Terminology. American National Standards Institute; New York: 1973.
- Caclin A, Brattico E, Tervaniemi M, Naatanen R, Morlet D, Giard MH, McAdams S. Separate neural processing of timbre dimensions in auditory sensory memory. *J Cogn Neurosci*. 2006; 18:1959–1972. [PubMed: 17129184]
- Dugundji J. Envelopes and pre-envelopes of real waveforms. *Information Theory IRE Trans*. 1958; 4:53–57.
- Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am*. 2001; 110:1150–1163. [PubMed: 11519582]
- Gfeller K, Christ A, Knutson JF, Witt S, Murray KT, Tyler RS. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *J Am Acad Audiol*. 2000; 11:390–406. [PubMed: 10976500]
- Gfeller K, Witt S, Woodworth G, Mehr MA, Knutson J. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann Otol Rhinol Laryngol*. 2002a; 111:349. [PubMed: 11991588]
- Gfeller K, Turner C, Mehr M, Woodworth G, Fearn R, Knutson JF, Witt S, Stordahl J. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int*. 2002b; 3:29–53. [PubMed: 18792110]
- Ghitza O. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *J Acoust Soc Am*. 2001; 110:1628–1640. [PubMed: 11572372]
- Gilbert G, Lorenzi C. The ability of listeners to use recovered envelope cues from speech fine structure. *J Acoust Soc Am*. 2006; 119:2438. [PubMed: 16642856]
- Grey JM. Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am*. 1977; 61:1270–1277. [PubMed: 560400]
- Gunawan D, Sen D. Spectral envelope sensitivity of musical instrument sounds. *J Acoust Soc Am*. 2008; 123:500–506. [PubMed: 18177177]
- Hong RS, Rubinstein JT. High-rate conditioning pulse trains in cochlear implants: dynamic range measures with sinusoidal stimuli. *J Acoust Soc Am*. 2003a; 114:3327–3342. [PubMed: 14714813]
- Hong RS, Rubinstein JT. Conditioning pulse trains in cochlear implants: effects on loudness growth. *Otol Neurotol*. 2006; 27:50. [PubMed: 16371847]
- Hong RS, Rubinstein JT, Wehner D, Horn D. Dynamic range enhancement for cochlear implants. *Otol Neurotol*. 2003b; 24:590. [PubMed: 12851550]
- Huss M, Moore BCJ. Dead regions and pitch perception. *J Acoust Soc Am*. 2005; 117:3841–3852. [PubMed: 16018486]
- Jolly CN, Spelman FA, Clopton BM. Quadrupolar stimulation for cochlear prostheses: modeling and experimental data. *IEEE Trans Biomed Eng*. 1996; 43:857–865. [PubMed: 9216159]

- Joris PX, Yin TC. Responses to the amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am.* 1992; 91:215–232. [PubMed: 1737873]
- Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear.* 2004; 25:173. [PubMed: 15064662]
- Krumhansl CL. Why is Musical Timbre so Hard to Understand. *Structure and Perception Electroacoustic Sound Music.* 1989:43–53.
- Laneau J, Wouters J, Moonen M. Improved music perception with explicit pitch coding in cochlear implants. *Audiol Neurotol.* 2006; 11:38–52.
- Licklider JCR, Pollack I. Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. *J Acoust Soc Am.* 1948; 20:42–51.
- Litvak LM, Delgutte B, Eddington DK. Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains. *J Acoust Soc Am.* 2003a; 114:2079–2098. [PubMed: 14587607]
- Litvak LM, Smith ZM, Delgutte B, Eddington DK. Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration. *J Acoust Soc Am.* 2003b; 114:2066–2078. [PubMed: 14587606]
- Liu S, Zeng FG. Temporal properties in clear speech perception. *J Acoust Soc Am.* 2006; 120:424–432. [PubMed: 16875238]
- Logan BF. Information in the zero crossings of bandpass signals. *Bell Syst Tech J.* 1977; 56:487–510.
- Marozeau J, de Cheveigné A, McAdams S, Winsberg S. The dependency of timbre on fundamental frequency. *J Acoust Soc Am.* 2003; 114:2946–2957. [PubMed: 14650028]
- McAdams S, Winsberg S, Donnadiou S, Soete G, Krimphoff J. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res.* 1995; 58:177–192. [PubMed: 8570786]
- McDermott HJ. Music perception with cochlear implants: a review. *Trends Amplif.* 2004; 8:49. [PubMed: 15497033]
- McDermott, HJ.; Looi, V. Perception of complex signals, including musical sounds, with cochlear implants. *Proceedings of the VIII International Cochlear Implant Conference; Indianapolis, IN.* 2004.
- Moore, BCJ.; Carlyon, RP. Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. Vol. 24. Springer; New York: 2005. p. 234
- Nimmons GL, Kang RS, Drennan WR, Longnion J, Ruffin C, Worman T, Yueh B, Rubinstein JT. Clinical assessment of music perception in cochlear implant listeners. *Otol Neurotol.* 2008; 29:149. [PubMed: 18309572]
- Nogueira W. A psychoacoustic “NofM”-type speech coding strategy for cochlear implants. *EURASIP J Appl Sign Proc.* 2005; 2005:3044–3059.
- Oxenham AJ, Bernstein JGW, Penagos H. Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci USA.* 2004; 101:1421–1425. [PubMed: 14718671]
- Papoulis A. Random modulation: a review. *IEEE Trans Acoustics, Speech Signal Processing.* 1983; 31:96–105.
- Plomp, R. *Hearing—Physiological Bases and Psychophysics.* Springer; Berlin: 1983. The Role of Modulation in Hearing; p. 270-276.
- Reiss LAJ, Turner CW, Erenberg SR, Gantz BJ. Changes in pitch with a cochlear implant over time. *J Assoc Res Otolaryngol.* 2007; 8:241–257. [PubMed: 17347777]
- Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil Trans Royal Soc London B.* 1992; 336:367–373.
- Ruffin, C.; Liu, G.; Drennan, W.; Won, JH.; Longnion, J.; Rubinstein, J. Evidence for Temporal Fine Structure Encoding by Cochlear Implant Subjects Using Envelope-Modulated Speech Processing Strategies. Poster Presented at the Association for Research in Otolaryngology Mid-Winter Meeting; 2007; 2007.
- Samson S, Zatorre RJ, Ramsay JO. Multidimensional scaling of synthetic musical timbre: perception of spectral and temporal characteristics. *Can J Exp Psychol.* 1997; 51:307–315. [PubMed: 9606948]

- Schimmel, S.; Atlas, L. Coherent envelope detection for modulation filtering of speech, acoustics, speech, and signal processing. 2005; Proceedings. (ICASSP '05). IEEE International Conference on; 2005. p. 221-224.
- Shamma S, Klein D. The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am.* 2000; 107:2631–2644. [PubMed: 10830385]
- Sheft S, Ardoint M, Lorenzi C. Speech identification based on temporal fine structure cues. *J Acoust Soc Am.* 2008; 124:562–575. [PubMed: 18646999]
- Shepherd RK, Javel E. Electrical stimulation of the auditory nerve: II. Effect of stimulus waveshape on single fibre response properties. *Hear Res.* 1999; 130:171–188. [PubMed: 10320107]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature.* 2002; 416:87–90. [PubMed: 11882898]
- Voelcker HB. Toward a unified theory of modulation. Part I: phase-envelope relationships. *Proc IEEE.* 1966; 54:340–353.
- Von Helmholtz, H.; Ellis, AJ. *On the Sensations of Tone as a Physiological Basis for the Theory of Music* Longmans, Green, and Co; 1895.
- Wilson BS. Engineering Design of Cochlear Implants. *Springer Handbook of Auditory Research.* 2004; 20:14–52.
- Xu L, Pfingst BE. Relative importance of temporal envelope and fine structure in lexical-tone perception. *J Acoust Soc Am.* 2003; 114:3024–3027. [PubMed: 14714781]
- Zeng FG, Nie K, Liu S, Stickney G, Del Rio E, Kong YY, Chen H. On the dichotomy in auditory perception between temporal envelope and fine structure cues. *J Acoust Soc Am.* 2004; 116:1351–1354. [PubMed: 15478399]

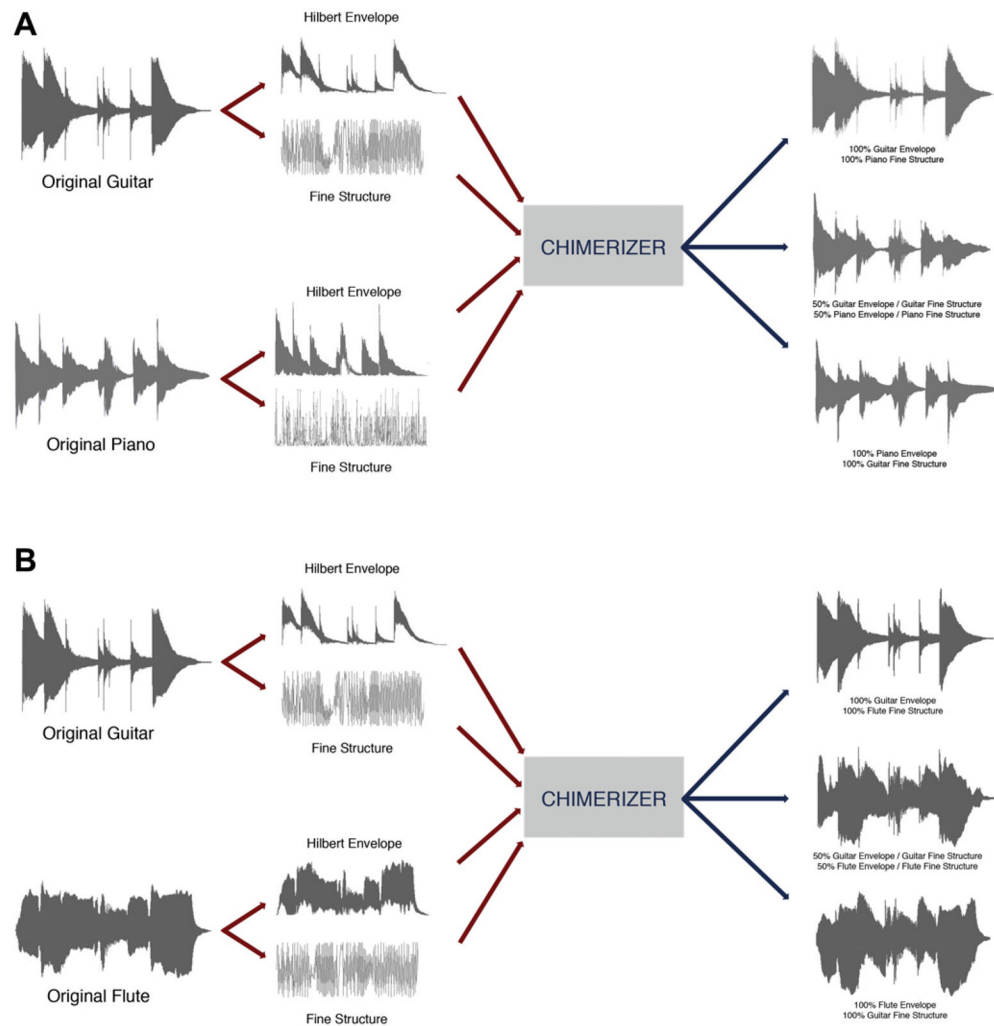


Fig. 1. Instrumental Chimera Synthesis. (A) represents chimera synthesis from two instruments with similar envelopes, and (B) represents chimera synthesis from two instruments with dissimilar envelopes. Two original instrument source files playing an identical melody are used as inputs into the chimerizer, which extracts the Hilbert envelope and fine structure of both signals. These extracted features are then recombined in variable ratios to produce a range of instrumental chimeras. Each sub-figure shows an example of three chimeras produced from two instruments. The three chimeras shown here represent one chimera composed of 100% instrument A envelope with 100% instrument B fine structure (top right), another composed of 50% instrument A envelope 50% instrument B envelope with 50% instrument A fine structure 50% instrument B fine structure (middle right), and a third composed of 100% instrument B envelope with 100% instrument A fine structure. For each instrumental pair, 25 chimeras were created, resulting in a total of 150 chimeras.

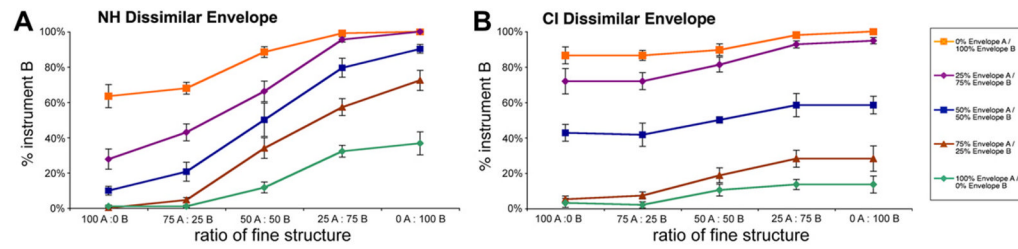


Fig. 2.

Comparison of timbre judgments of NH subjects ($n = 14$) and CI Subjects ($n = 12$) for instrumental chimeras synthesized from source instruments with dissimilar envelopes. Standard error bars are shown. The y -axis represents the percentage of times the subject identified the given chimera as sounding most similar to instrument B, and the x -axis represents the ratio of fine structure for instruments A and B in the chimera.

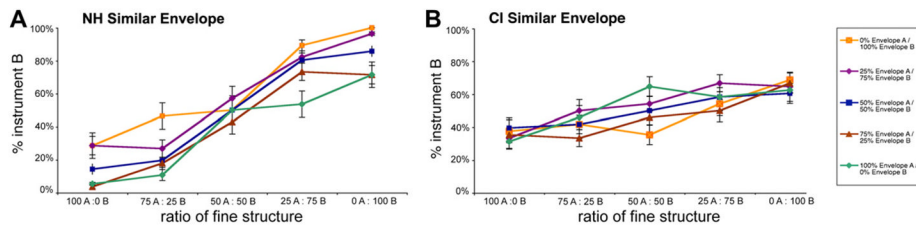


Fig. 3. Comparison of timbre judgments of NH subjects ($n = 14$) and CI subjects ($n = 12$) for instrumental chimeras synthesized from source instruments with similar envelopes. Standard error bars are shown. The y -axis represents the percentage of times the subject identified the given chimera as sounding most similar to instrument B, and the x -axis represents the ratio of fine structure of instruments A and B in the chimera.

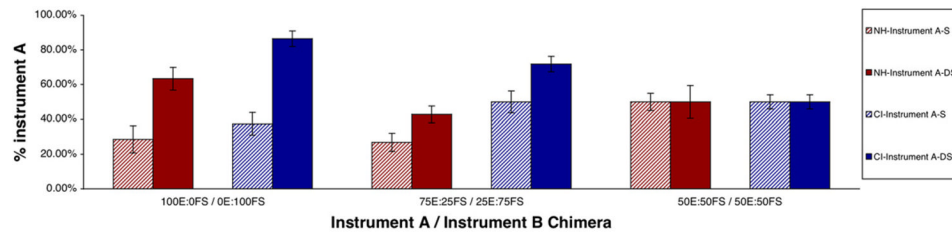


Fig. 4. Head-to-Head comparison of timbre identification of ambiguous chimeras between NH subjects and CI users for similar (S) and dissimilar (DS) envelope pairs. Each histogram summarizes the mean and standard error of responses to ambiguous chimera stimuli. The y -axis represents the percentage of times the subject identified instrument A in the chimera presented, and the x -axis represents the ratio of both envelope and fine structure of the two source instruments used to generate the instrument chimera.

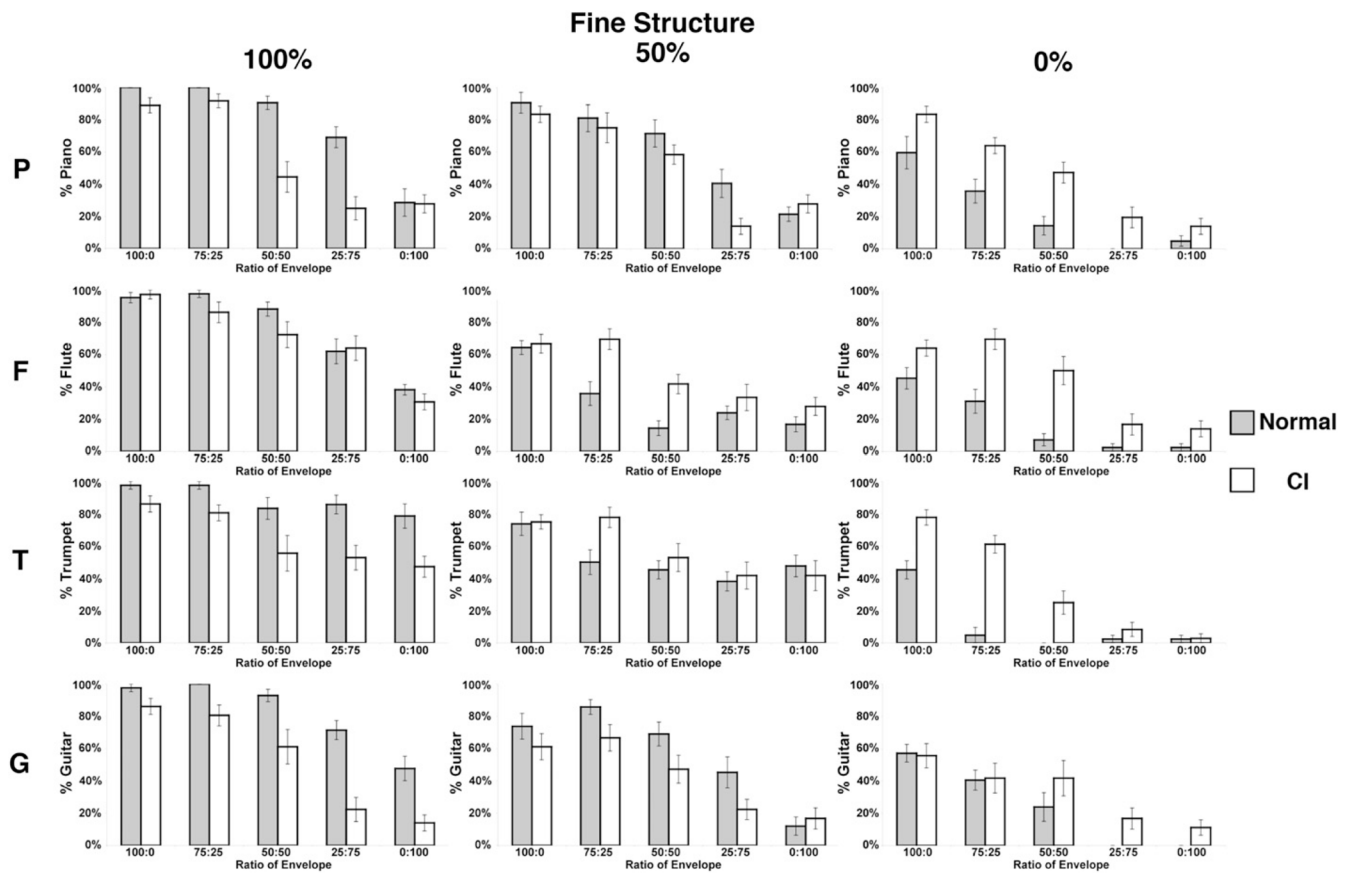


Fig. 5. Comparison of timbre judgments of all subjects ($n = 26$) for all instrumental chimeras. Each histogram summarizes mean and standard error of timbre judgments. Responses are categorized by instrument and fine structure of indicated instrument. The y-axis represents the percentage of times the subject identifies the indicated instrument in the given chimera, and the x-axis represents the ratio of envelope of the two source instruments.

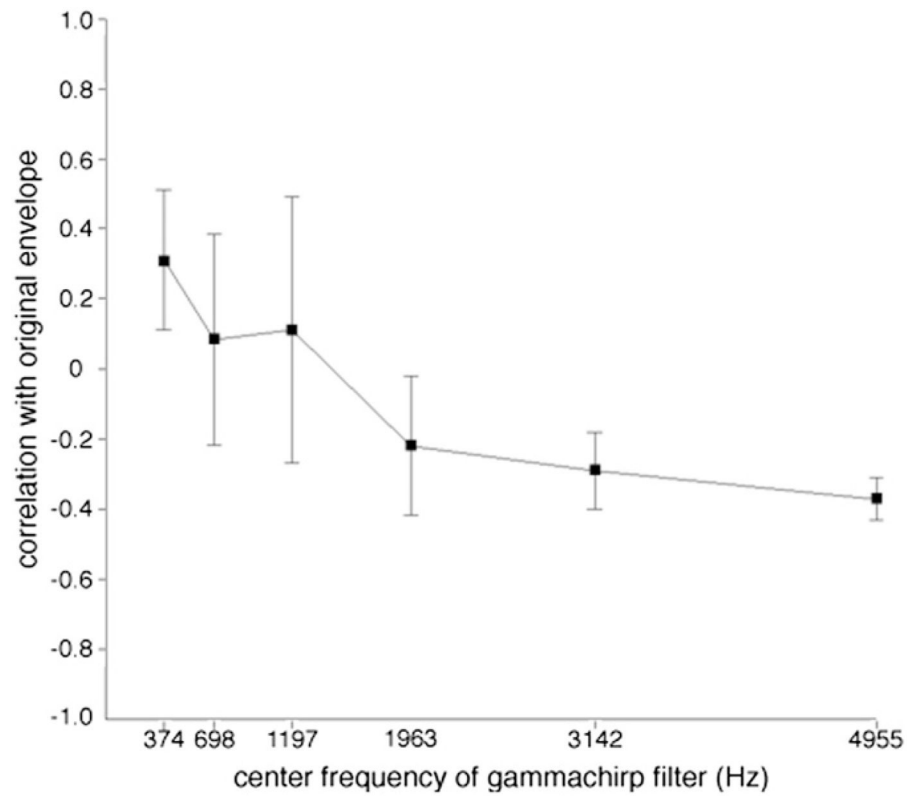


Fig. 6. Correlation between original and recovered envelopes of instrumental chimeras. Mean correlation coefficients, with standard deviations, were computed (across 150 instrumental chimeras) between the original envelopes and the recovered envelopes of the instrumental chimeras at the output of six gammachirp filters. The y -axis represents the correlation coefficient between the original and recovered envelopes, and the x -axis indicates the center frequencies of the gammachirp auditory filters used.

Table 1

Demographic data for cochlear implant users.

Subject	Gender	Age at Testing (Yrs.)	Length of Implant (Mo.)	Years of PHL	Implant Type	Strategy	HINT-Q %	HINT-N %	Musical Experience (Yrs)
CI1	F	30	119	4	ABC CII	Hi-Res P	98	100	-
CI2	M	29	14	1	ABC Hi-Res 90K	Hi-Res P	57	-	13 (Guitar)
CI3	F	52	71	1	CC N24	ACE	97	100	-
CI4	F	59	36	2	CC NF	ACE	98	87	-
CI5 (B)	F	55	84	2	CC N24	ACE	98	87	-
CI6	M	72	79	<1	ABC CII	Hi-Res P	93	-	15 (Piano)
CI7	F	65	12	<1	ABC Hi-Res 90K	Hi-Res P	91	83	-
CI8	M	44	120	2	ME Combi	CIS	98	52	-
CI9	F	57	27	5	CC NF	ACE	98	55	-
CI10 (B)	M	48	132	15	ABC Clarion	CIS	95	87	14 (Piano)
CI11	F	64	99	20	ABC CII	SAS	96	94	10 (Piano)
CI12	M	64	28	2	ABC Hi-Res 90K	Hi-Res P	96	35	6 (Piano)
Mean		53.3	68.4						
SD		13.5	43.8						

The demographic data of the cochlear implant population is represented, with data on their age at testing, length of implant usage, years of profound hearing loss before implantation (PHL), devices, implant-based speech processing strategies, HINT (Hearing-In-Noise Test) scores and years of musical experience provided. Dash marks are given where test was not performed. (B) = Bilateral CI recipient, ABC = Advanced Bionics Corporation, CC = Cochlear Corporation, ME = Med-El, N24 = Nucleus 24, CII = Clarion II, NF = Nucleus Freedom, Hi-Res P=Hi-Resolution Paired, ACE = Advanced Combination Encoders, CIS = Continuous Interleaved Sampling, SAS = Simultaneous Analog Stimulation.