Published in final edited form as:

Ear Hear. 2014 ; 35(3): 379–382. doi:10.1097/AUD.000000000000013.

Melody recognition in dichotic listening with or without frequency-place mismatch

Ning Zhou^{a)} and

Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina 27834

Li Xu

School of Rehabilitation and Communication Sciences, Ohio University, Athens, Ohio 45701

Abstract

Objective—The purpose of the study was to examine recognition of degraded melodic stimuli in dichotic listening with or without frequency-place mismatch.

Design—Melodic stimuli were noise-vocoded with various number-of-channel conditions in a dichotic and monaural processor. In the dichotic zipper processor, the odd-indexed channels were tonotopically matched and presented to the left ear while the even-indexed channels were tonotopically matched or upward shifted in frequency and presented to the right ear. In the monaural processor, all channels either unshifted or shifted were presented to the left ear alone. Familiar melody recognition was measured in 16 normal-hearing adult listeners.

Results—Performance for dichotically presented melodic stimuli did not differ from that for monaurally presented stimuli even with low spectral resolution (8 channels). With spectral shift introduced in one ear, melody recognition decreased with increasing spectral shift in a non-monotonic fashion. With spectral shift, melody recognition in dichotic listening was either not different or superior in a few cases relative to the monaural condition.

Conclusions—With no spectral shift, cohesive fusion of dichotically presented melodic stimuli did not seem to depend on spectral resolution. In spectrally shifted conditions, listeners may have suppressed the partial shifted channels in the right ear and selectively attended only to the unshifted ones, resulting in dichotic advantages for melody recognition in some cases.

Keywords

melody recognition; dichotic listening; spectral shift

1. Introduction

Dichotic listening involves simultaneously presenting different auditory signals to the two ears. The inputs to the two ears can be two competing signals or complementary components (e.g., formant frequencies) split from one auditory signal (e.g., Broadbent and Ladefoged,

^{a)} Author to whom correspondence should be addressed: Ning Zhou, Ph.D., Department of Communication Sciences and Disorders, East Carolina University, Greenville, NC 27834, zhoun@ecu.edu.

1957). The ability of the central auditory system to integrate complementary components from the two ears is known as “spectral fusion”, which can be affected by relative onset time, relative intensity, or the fundamental frequencies of the split components (Cutting, 1976). Recent studies have shown that spectrally degraded speech signals cannot be integrated in dichotic listening, if spectral resolution of the original speech is so low that the split components cannot be recognized as coming from the same vocal source (e.g., Loizou et al., 2003). In these vocoder studies (Dorman et al., 2001; Loizou et al., 2003), a dichotic zipper processor was used where the two ears were alternately stimulated with the even-indexed spectral channels assigned to one ear and the odd-indexed channels assigned to the other ear. Listeners’ speech recognition with fewer than 12 channels declined significantly from the monaural condition when the channels were split and presented dichotically (Loizou et al., 2003).

The first aim of the study is to investigate whether spectral resolution is also required for fusing dichotically presented melodic stimuli. A fundamental difference between music and speech perception is that music perception relies heavily on spectral or temporal fine structure cues for resolving harmonic structures for place pitch (Smith et al., 2002), while temporal envelopes carried by a small number of spectral channels are sufficient for understanding speech in quiet (Shannon et al., 1995). The number of channels required for good music perception probably exceeds 30 (Kong, et al., 2004), far greater than the functionally available channels provided in modern cochlear implant devices. Given the reliance of music perception on spectral information, listeners might demand equal, if not higher, spectral resolution to integrate dichotically presented melodic stimuli compared to speech stimuli.

Frequency-place mismatch occurs when the analysis bands are excited at mismatched tonotopic places by shifting the carrier bands, typically to higher frequencies. This is an acoustic simulation of a shallowly or partially inserted cochlear implant (e.g., Dorman et al., 1997; Fu and Shannon, 1999; Zhou and Xu, 2008). Speech recognition has been found to significantly deteriorate with this type of distorted spectral cues (e.g., Dorman et al., 1997; Fu and Shannon, 1999; Zhou et al., 2010), because accurate place representation of formant frequencies appears to be important for perceiving vowels and place of articulation of consonants. The same effects were found for split channels of speech stimuli presented dichotically with frequency-place shift only occurring in one ear (Siciliano et al., 2010).

The second aim of the study is to determine whether music perception would be affected by shifting partial channels in one ear in a similar fashion as speech recognition. When place pitch is not fully conveyed as is the case with electric hearing or its acoustic simulation, listeners would also make use the temporal envelope cues besides spectral cues for pitch perception, as was demonstrated for lexical tone perception (Xu et al., 2002; Zhou and Xu, 2008). Unlike speech recognition where the envelope information in each channel is unique, thus all channels must integrate for a complete spectrum, we hypothesized that for music perception listeners might be able to attend only to the unshifted partial channels of the split melodic stimuli to a certain degree, making them less susceptible to the spectral distortion. In this experiment, we compared a dichotic condition where channels were basally shifted

only in one ear to a monaural condition where the shifted channels were summed with the unshifted channels and presented to a single ear.

The results of the study have clinical implications for cochlear implant users who are bilaterally implanted. The results will provide insights into pitch matching between bilaterally implanted devices fitted with dichotic clinical maps.

2. Methods

Sixteen normal-hearing subjects with ages ranging from 22 to 28 years old were recruited. The subjects had not previously received formal music training but were required to know the 10 melodies used in the music tests. The use of human subjects was approved by the Ohio University Institutional Review Board.

The stimulus materials were comprised of 10 familiar melodies (e.g., Happy Birthday, Twinkle, Twinkle, Little Star, etc.), each consisting of 10 digitally synthesized piano tones. Each of the piano tones was 300 ms in duration. All piano tones were processed to have identical temporal and spectral envelopes (see Nimmons et al., 2008 for details). The fundamental frequency range of tones was 196 Hz to 533 Hz (i.e., G3 - C5). The stimuli were subject to a noise-excited vocoder using a dichotic and monaural processor in which tonotopically matched or frequency-place mismatched conditions were created. The stimuli were first bandpass filtered into 8, 12, 16, and 20 spectral channels in the dichotic processor and 8 and 20 channels in the monaural processor, both covering a frequency range of 269–3282 Hz, based on the Greenwood (1990) formula. The spectral channels were odd or even indexed. The output of each band was half-wave rectified and low-pass filtered at 400 Hz (2nd-order Butterworth, 12 dB/octave) to extract the temporal envelopes. The extracted envelope was then used to modulate a wide-band noise. The modulated signals were band-restricted using a bank of bandpass filters, the frequencies of which (i.e., carrier bands) were systematically varied to create tonotopically-matched or frequency shifted conditions. Under all conditions, the carrier bands of the odd-indexed channels were always tonotopically-matched to the corresponding analysis bands. The carrier bands of the even-indexed channels were shifted basally from the analysis bands by 0, 2, 3, 4, 5, 6, and 7 mm in tonotopic distance. In the dichotic zipper processor (Fig. 1, left), the odd-indexed unshifted channels were summed and presented to the left ear, and the even-indexed shifted channels were summed and presented to the right ear. In the monaural processor (Fig. 1, right), the unshifted and shifted channels were summed and presented to the left ear alone.

The perceptual task was a familiar melody recognition test that used the 10 vocoded melodies. No practice or training was provided. During the test, vocoded stimuli were fully randomized and delivered at the most comfortable level through headphones (Sennheiser HD 265) in a sound booth. All 16 subjects completed the test using the dichotic processor and 11 of them continued with the test using the monaural processor.

3. Results

Familiar melody recognition performance is shown as a function of the amount of basal shift of the even-indexed channels for the dichotic processor (Fig. 2, left panel) and monaural

processor (Fig. 2, filled symbols in right panel). For the dichotic processor, ANOVAs showed that both mismatch [$F(6, 45) = 22.03, p < 0.001$] and number of channels [$F(3, 18) = 81.76, p < 0.001$] were significant factors. Interaction between number of channels and mismatch was also significant [$F(18, 270) = 2.66, p < 0.001$]. For the monaural processor, both mismatch [$F(6, 10) = 7.83, p < 0.001$] and number of channels [$F(1, 6) = 65.99, p < 0.001$] were also significant factors. Interaction between the two factors was significant [$F(6, 60) = 2.40, p = 0.03$].

For both dichotic and monaural processors, although scores in each shifted condition were significantly worse than the unshifted conditions ($p < 0.0083$ in all cases after Bonferroni correction), performance was reduced with increasing mismatch in a non-monotonic fashion. For both processors, scores for the 4-mm-shift condition were significantly better than those for the 2- and 3-mm shift conditions ($p < 0.025$ in all cases after Bonferroni correction). Subjective reports from the listeners did not suggest evidence of hearing two auditory images in any of the spectral shifted conditions for dichotic listening.

The right panel of Fig. 2 shows data taken from the 11 subjects who completed both the dichotic and monaural conditions. Two-way repeated-measures ANOVAs with processor and mismatch as two factors showed that performance of the two processors was not significantly different. For both the 8-channel and 20-channel conditions, the factor of processor was not significant [8 channels: $F(1, 10) = 1.38, p = 0.26$, 20 channels: $F(1, 10) = 1.31, p = 0.28$; estimated power = 7.59%], and mismatch remained a significant factor [8 channels: $F(6, 60) = 11.37, p < 0.001$; 20 channels: $F(6, 60) = 15.68, p < 0.001$]. However, the interaction between processor and mismatch was significant for the 8-channel condition [$F(6, 60) = 4.93, p < 0.05$], although not significant for the 20-channel condition [$F(6, 60) = 1.93, p = 0.26$]. Given the probability of detecting the effect of processor was low in the ANOVA analysis as revealed by the power analysis, paired comparisons were conducted for comparing the two processors for each amount of mismatch. The 8-channel dichotic processor was significantly better than the monaural processor with 2 mm shift [$t(10) = 2.78, p = 0.02$] and the 20-channel dichotic processor was better than the monaural processor with 4 mm shift [$t(10) = 3.73, p = 0.003$].

4. Discussion and conclusions

There is no strong evidence to indicate that integration of dichotically presented musical stimuli was different with higher spectral resolution than with lower spectral resolution. Melody recognition decreased with frequency-place mismatch introduced in one ear in a non-monotonic fashion. With spectral shift, melody recognition in dichotic listening was either comparable or superior in some cases relative to monaural listening.

In unshifted conditions, melody recognition in dichotic listening was not different from that in monaural listening indicating cohesive integration for both the 8- and 20-channel conditions. Loizou et al. (2003) indicated that a sufficiently high spectral resolution (12 channels) is required for sentence and vowel stimuli to be recognized as belonging to the same vocal source when the channels are split. Our results showed that melodic stimuli were cohesively integrated even though there were only as few as 4 channels presented in each

ear. The reason could be that since the stimuli were not speech produced by characterized vocal structures, spectral fusion would not have to depend on recognizing the source of the split components. Therefore, although performance did differ between 20 and 8 channels, which confirms the importance of spectral resolution for music perception, splitting these channels in two ears did not affect the performance for the low number-of-channel condition. It is also possible that the effects may be found for more challenging music perception tasks such as pitch contour identification.

Frequency-place mismatch seemed to affect melody recognition in a different fashion than what was found for speech recognition. First of all, melody recognition performance decreased with increasing spectral in a non-monotonic fashion for both processors. The non-monotonic decrease was seen with a rebound in performance at the 4-mm shift. With a 4-mm shift, carrier bands of the even-indexed channels were shifted up to approximately an octave higher than the unshifted condition (Fig. 2, top axis). The frequencies of the carrier bands were in a rough harmonic relationship with the frequencies prior to the shift. This condition may have introduced a smaller spectral distortion than the other shifted conditions, and may explain the increased performance with that particular amount of shift.

Secondly, performance with partially shifted channels was significantly better in some cases than performance of the monaural condition where all channels, shifted or unshifted, were mixed and presented to one ear. In the manipulation of spectral shift, only the even-indexed channels were shifted upward in frequencies, which means frequency-place mismatch was present only in one ear in dichotic listening. One explanation for the listeners' performance is that they suppressed or ignored mismatched information and attended to information in the unshifted channels. Therefore, performance of the dichotic processor was better compared to that of the monaural processor where the mix of the unshifted with the shifted channels was likely to introduce interference. Siciliano et al. (2010) reasoned that suppression of the mismatched channels is not likely for *speech* perception since all channels carry unique spectral information necessary for restoring a complete spectral envelope. Nonetheless, suppression of the distorted pitch information may be possible for melody recognition assuming that shifted channels of a music note may not provide substantially more pitch information in addition to the unshifted channels. These speculations however would require testing the partial channels in the left ear alone to confirm.

In summary, results of the present study indicate that spectral resolution has minimal effect on integration of musical stimuli in dichotic listening. Without spectral mismatch, melody recognition in dichotic listening was comparable to that in monaural listening. In spectrally shifted conditions, melody recognition in dichotic listening was either comparable or superior to that in monaural listening. These results have clinical implications for pitch matching electrode arrays with different insertion depths in dichotic mapping.

Acknowledgments

The study was supported in part by the NIDCD/NIH grants R15 DC009504 and T32-DC00011.

References

- Broadbent DE, Ladefoged P. On the fusion of sounds reaching different sense organs. *J Acoust Soc Am.* 1957; 29:708–710.
- Cutting JE. Auditory and linguistic process in speech perception: Inferences from six fusions in dichotic listening. *Psychol Rev.* 1976; 83:114–140. [PubMed: 769016]
- Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear implant electrode insertion depth on speech understating. *J Acoust Soc Am.* 1997; 102:2993–2996. [PubMed: 9373986]
- Dorman, M.; Loizou, P.; Spahr, AJ.; Maloff, ES.; Wie, SV. Speech understanding with dichotic presentation of channels: Results from acoustic models of bilateral cochlear implants. Conference on Implantable Auditory Prostheses; Asilomar, Monterey, CA. 2001.
- Fu QJ, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J Acoust Soc Am.* 1999; 105:1889–1990. [PubMed: 10089611]
- Greenwood DD. A cochlear frequency-position function for several species-29 years later. *J Acoust Soc Am.* 1990; 87:2592–2605. [PubMed: 2373794]
- Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hear.* 2004; 25:173–185.
- Loizou PC, Mani A, Dorman MF. Dichotic speech recognition in noise using reduced spectral cues. *J Acoust Soc Am.* 2003; 114:475–483. [PubMed: 12880058]
- Nimmons GL, Kang RS, Drennan WR, et al. Clinical assessment of music perception in cochlear implant listeners. *Otol Neurotol.* 2008; 29:149–155. [PubMed: 18309572]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304. [PubMed: 7569981]
- Siciliano CM, Faulkner A, Rosen S, et al. Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *J Acoust Soc Am.* 2010; 127:1645–1660. [PubMed: 20329863]
- Smith Z, Delgutte B, Oxenham A. Chimaeric sounds reveal dichotomies in auditory perception. *Nature.* 2002; 416:87–90. [PubMed: 11882898]
- Xu L, Tsai Y, Pfingst BE. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J Acoust Soc Am.* 2002; 112:247–258. [PubMed: 12141350]
- Zhou N, Xu L. Lexical tone recognition with spectrally mismatched envelopes. *Hear Res.* 2008; 246:36–43. [PubMed: 18848614]
- Zhou N, Xu L, Lee C-Y. The effects of frequency-place mismatch on consonant confusion. *J Acoust Soc Am.* 2010; 128:401–409. [PubMed: 20649234]

Figure 1.

Schematic representation of the dichotic and monaural processors. Left panel: in the dichotic processor, the unshifted odd-index channels are delivered to the left ear and the basally shifted even-index channels are delivered to the right ear. Right panel: in the monaural processor, both the unshifted odd-index channels and the basally shifted even-index channels are delivered to the left ear alone.

Figure 2.

Familiar melody recognition performance. Left panel: Familiar melody recognition performance for the dichotic processor is plotted as a function of spectral shift in the even-index channels presented in the right ear. Right panel: Familiar melody recognition scores from the 11 subjects who completed both dichotic (dashed lines) and monaural (solid lines) testing are shown as a function of spectral shift in the even-index channels. Spectral shift in tonotopic distance (mm) is shown on the bottom abscissa and spectral shift in frequency (octave) is shown on the top abscissa.