Lexical Tone Perception in Musicians and Non-musicians

Jennifer A. Alexander[1], Patrick C.M. Wong[2,3], and Ann R. Bradlow[1]

[1]Department of Linguistics
[2]Department of Communication Sciences and Disorders
[3]Northwestern University Institute for Neuroscience
Northwestern University, Evanston, IL U.S.A.
jenalex@northwestern.edu

Abstract

It has been suggested that music and speech maintain entirely dissociable mental processing systems. The current study, however, provides evidence that there is an overlap in the processing of certain shared aspects of the two. This study focuses on fundamental frequency (pitch), which is an essential component of melodic units in music and lexical and/or intonational units in speech. We hypothesize that extensive experience with the processing of musical pitch can transfer to a lexical pitch-processing domain. To that end, we asked nine English-speaking musicians and nine English-speaking non-musicians to identify and discriminate the four lexical tones of Mandarin Chinese. The subjects performed significantly differently on both tasks; the musicians identified the tones with 89% accuracy and discriminated them with 87% accuracy, while the non-musicians identified them with only 69% accuracy and discriminated them with 71% accuracy. These results provide counter-evidence to the theory of dissociation between music and speech processing.

1. Introduction

Music and speech have much in common. For one thing, they represent the two most cognitively complex uses of sound-based communication across any of the world's species. In addition, both speech and music are generative: a finite number of simple elements such as pitches or segments combine hierarchically ("syntactically") to create increasingly complex, meaningful structures, like words and utterances, melodies and songs [1]. In view of these functional similarities, a considerable body of previous work has investigated the extent to which music and speech share common processing mechanisms. One possibility, which assumes a strict interpretation of modularity, is that music and speech are cognitively unique and distinct in that they maintain discrete mental processing systems [2]. A number of behavioral and imaging studies produced within this framework have given rise to the idea of hemispheric lateralization or dominance, which suggests that linguistic processing takes place within the left hemisphere of the brain, while music processing occurs in the right hemisphere [3-4]. An alternative possibility is that hemispheric dominance pertains to particular aspects of auditory processing and that shared acoustic features of speech and music will be processed similarly. In support of this alternative several studies have shown that the left hemisphere tends to handle phonemic processing – the processing of words, syllables, and lexical tones – while the right processes melodic and prosodic units, like musical phrases, intonational phrases, pitch contours, and affect (see, e.g., [5-14]).

Moreover, the picture of hemispheric dominance for separate aspects of processing is not as absolute as it may seem. As Wang *et al* [14] note, the lateralization of the brain is but a tendency; "dominance" does not always exclude activity in the other hemisphere. Recent behavioral and neural studies have shown that the lateralization boundaries can in fact be blurred. Certain shared aspects of music and speech, such as hierarchical ("syntactic") organization, appear to be processed in overlapping areas, suggesting that there are common neural mechanisms subserving speech and music [15]. If this is the case, then it would not be surprising to see behavioral manifestations of sharing between music and speech.

The current study seeks to identify another such similarity in music and speech processing by directly investigating the effect of experience-dependent learning in one domain (in this case, music) on processing in the other domain (in this case, speech). Specifically, we hypothesize that there is an overlap in the processing of fundamental frequency (pitch) in these two domains such that extensive experience with pitch processing in music will be manifested as "enhanced" pitch processing in speech. Systematic pitch variation is a fundamental feature of both music and speech; music incorporates pitch changes within a specified parameter (the "key" of the piece) to express compositional and affective meaning, while in speech, pitch is used to convey pragmatic information and, in the case of tone languages, lexical information (i.e., contrasts in word meaning can be conveyed via pitch pattern alone). In this study, American English-speaking musicians and non-musicians with no tone language experience were asked to identify and discriminate the four lexical tones of Mandarin Chinese (high-level, high-rising, low-dipping, and low-falling). Results suggest that extensive experience with musical pitch processing may facilitate lexical pitch processing in a novel tone language to a significant degree.

2. Methods

2.1. Subjects

Three groups of subjects participated in this study. The first was comprised of five adult female native speakers of Mandarin Chinese who also speak English ("Mandarin speakers"). All the subjects in this group ranked Mandarin as their dominant language relative to English and spoke mostly Mandarin in their childhood. The Mandarin speakers ranged from 24 to 36 years of age, with a mean age of 29.8 years and SD of 5.4 years. The second group consisted of seven adult female and two adult male native speakers of

American English who had no previous exposure to Mandarin and who had eight or more years of continuous private piano or voice lessons up until or beyond the year 2002 ("American musicians"). The American musicians ranged from 18 to 26 years of age, with a mean age of 20.2 years and SD of 2.4 years. Of the American musicians, six considered their primary instrument to be voice, and three considered their primary instrument to be piano. The third group was composed of five adult female and four adult male monolingual native speakers of American English who had no previous exposure to Mandarin, no more than three years of continuous private music lessons of any sort, and had not studied any instrument beyond the year 1997 ("American non-musicians"). The American non-musicians ranged in age from 18 to 34 years, with a mean of 22.3 years and SD of 4.9 years. Groupings were based on participants' answers to a detailed music- and language-experience questionnaire. None of the subjects had any known hearing problems, and none had a cold or ear infection on testing day.

2.2. Stimuli

Stimuli consisted of twenty monosyllabic Mandarin Chinese words. The five syllables *bu, di, lu, ma, mi* were each produced in citation form with the four tones of Mandarin. These particular syllables were chosen because they were comprised of segments found in American English as well. Therefore, American listeners would presumably find it easier to focus their attention on the tones. Talkers consisted of two male and two female native speakers of Mandarin Chinese, all of whom were in their twenties and were from Beijing. Using PRAAT [16], the stimuli were sampled at 44.1 kHz, normalized for RMS amplitude at 70 dB, and normalized for average duration at 0.442 seconds. Normalization for duration was a critical step, as duration can itself be a cue for tone perception [17]. The citation form of the dipping tone is generally longer in duration than the other tones; pilot studies found that listeners were able to disregard pitch in favor of duration in order to identify and discriminate the third tone. Without the duration difference, pitch became the most salient perceptual cue. Five native Mandarin-speaking informants transcribed the stimuli for intelligibility, writing the phonetic transcription (using the pin-yin transcription convention) of the word they heard. They transcribed the words with a mean accuracy rate of 99.8%.

2.3. Procedures

There were two experiments in this study: tone identification and tone discrimination. Both experiments used the same participants and stimuli; both took place within one testing session. All experiments were presented via E-PRIME software [18]. The order of experiments was counterbalanced across participants, and all stimuli within an experiment were presented in random order.

Before the start of the first experiment, all subjects were given a short tutorial, the purpose of which was to familiarize the subjects with the lexical tone system of Mandarin. They were first informed that all stimuli consisted of real Mandarin Chinese words. They then learned about the four lexical tones and learned how to differentiate them according to their pitch patterns. Finally, they listened to four words, each of which exemplified a different tone as produced by a different talker. The tutorial was repeated as many times as the subject desired; most chose to repeat it twice.

2.3.1. Experiment 1: tone identification

The first experiment was a forced-choice identification task. Each tonal pattern was represented visually on the computer screen by a picture of an arrow. For example, the high-level tone was represented by a horizontal arrow that pointed straight to the right; the rising tone corresponded to an arrow pointing upward. In a given trial, participants simultaneously heard one word and saw two arrows side-by-side on the computer screen; exactly one of the arrows matched the tone in the word that they heard. Participants indicated which of the two arrows ("A" or "B") corresponded to the tone in the word they heard by pressing the appropriately-marked button on a button box. Subjects were encouraged to respond as quickly and accurately as possible, not sacrificing one for the other, but the inter-stimulus-interval was set at 3 seconds. Both accuracy and reaction time were logged.

2.3.2. Experiment 2: tone discrimination

The second experiment was a forced-choice discrimination (AX) task. For this task, the words were arranged in pairs. Subjects were asked to indicate via button-press whether the tones in the word pairs were the same ("A") or different ("B"). Again, subjects were encouraged to respond as quickly and accurately as possible, and the inter-stimulus-interval was set at 3 seconds. Both accuracy and reaction time were logged in this experiment as well.

3. Results

On average, the Mandarin speakers correctly identified the tones 97% of the time (range = 94%-99%) and accurately discriminated them 89% of the time (range = 80%-95%). The American musicians correctly identified the tones 89% of the time (range = 76%-96%) and discriminated them 87% of the time (range = 75%-96%). The American non-musicians correctly identified the tones only 69% of the time (range = 60%-79%) and discriminated them 71% of the time (range = 63%-87%). The musicians' and non-musicians' data were submitted to one-way ANOVAs with musical experience as the factor. The analysis showed a significant difference between the musicians and non-musicians groups in terms of accuracy of performance on both identification [$F(1, 16) = 28.828$, $p < 0.001$] and discrimination [$F(1, 16) = 21.84$, $p < 0.001$]. In addition, reaction-time data showed no speed-accuracy trade-off, as the musicians were both faster and more accurate than the non-musicians. Mean reaction-time for the musicians on the identification task was 1286.11 ms, as compared with 1464.48 ms for the non-musicians (and 1057.13 ms for the Mandarin speakers). These results are summarized in figures 1 and 2. The Mandarin speakers' data have been included for purposes of visual comparison. They provided a means for assessing the feasibility of the tasks.

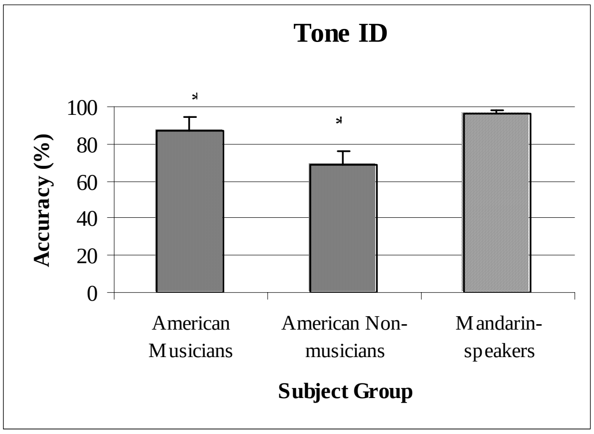

Figure 1: American musician and non-musician responses on the tone identification task. Error bars refer to standard deviation.

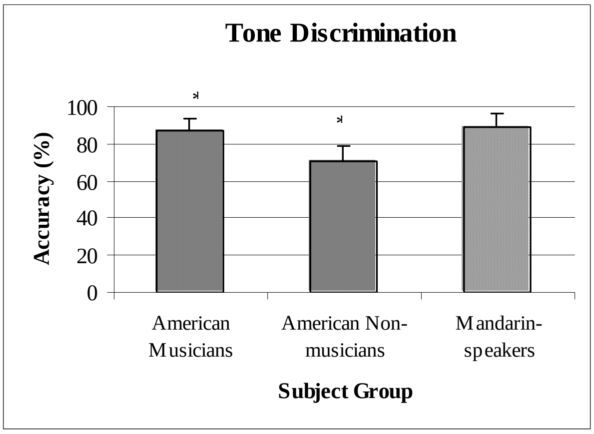

Figure 2: American musician and non-musician responses on the tone discrimination task. Error bars refer to standard deviation.

4. Discussion

One of the more obvious parallels between music and speech is that both utilize manipulations in pitch to convey meaning. Despite this and other similarities, many have thought them to maintain discrete mental processing systems. Several studies produced within this framework have suggested that, when it comes to pitch processing, the left hemisphere handles lexical tones while the right processes melodic (and prosodic) units. The current paper, along with some recently published abstracts by Gottfried and colleagues [19-20], suggests instead that musicians may be able to transfer their music pitch-processing abilities to a speech pitch-processing task. Having made these claims, it should be noted that there exists another possible interpretation of our data: it may actually be the case that the subjects in the “musicians” group performed better than the non-musicians *not* because they had musical experience *per se*, but rather because they possessed a higher pitch-processing aptitude than the non-musicians.

Our findings provide an interesting parallel to recent investigations of other aspects of speech and music processing. For example, recent neural studies by Patel and colleagues suggest that there is a possible overlap in the processing of linguistic and musical syntax. Musical syntactic processing has been found to activate both Broca’s and Wernicke’s language areas, which are located in the left hemisphere; likewise, P600 event-related potential (ERP) data have shown that the right hemisphere is activated during linguistic syntax processing in musicians [15]. These findings contrast with neuropsychological evidence that linguistic and musical syntax are dissociable (see, among others, [21]). The behavioral evidence found in the current study complements these existing neural studies and suggests another way in which music and speech might overlap.

5. Future work

This study lends itself to a number of interesting lines of future work. First, we might want to explore the possibility that different types of music experience could potentially influence the lexical pitch-processing mechanism in different ways. To be explicit, it might be the case that the music-pitch processing abilities acquired via study of one particular instrument might transfer easily to a lexical pitch processing domain, whereas those acquired via study of a different instrument might not. This hypothesis has been borne out in a newly-begun pilot study that compares the lexical pitch processing abilities of pianists and vocalists. Preliminary data suggest that pianists respond less accurately and more slowly than vocalists to the same tone perception experiments used in the current study. Intuition tells us that a cause of the disparity might be that vocalists, unlike pianists, have experience manipulating ever-changing musical pitch excursions, and that they are able to extend this experience to the lexical pitch processing domain. It stands to reason that pianists, who manipulate only discrete pitches on a keyboard, might not fare so well on such experiments.

The effects of one’s music training might extend to his production of lexical tone as well. For example, another obvious point of comparison between vocalists and pianists is that vocalists utilize their vocal mechanism to manipulate pitch and pianists do not. The inclusion of a lexical tone imitation/repetition task might uncover another way in which the lexical tone pitch processing skills of these two different types of musicians differ. If, for instance, vocalists but not pianists were found to possess the ability to accurately imitate lexical tones, it would suggest that experience with vocal manipulation of music pitch is also transferable to a lexical pitch manipulation domain.

A different line of work might investigate the flip side of the current study. Since our data suggest that music pitch processing experience can transfer to a lexical pitch perception domain, we might wonder whether the reverse is true as well. In other words, can pitch perception skills honed during long-term, continuous, tone language experience transfer to the musical-pitch processing domain? As yet, no known precedence has been established for investigation of this question. One might begin by creating experiments that parallel those used in the current study; this would set the course for a meaningful comparison.

Finally, one might investigate the connection between the current study, which focuses on the perception of tones stripped of semantic content, and the learning of meaningful tone words that vary only in their tone. A recent study by Wong and Perrachione [22] suggests that a heightened ability to recognize pitch patterns in non-lexical contexts was closely related to the subjects' speed and success in a tone word-learning task. That is, people who easily identified non-linguistic pitch patterns were more successful at learning meaningful words with varying tones than people who had trouble recognizing lexical pitch patterns. This, taken with our finding that there is a connection between music pitch processing and lexical pitch processing, raises the possibility that musicians would be more adept at learning meaningful words in a tone language than non-musicians.

6. Conclusions

Many consider music and speech to have separate cognitive systems; however, the current study found evidence that there is overlap in the processing of pitch, an aspect of sound that is common to both. In a set of two perception experiments, American-English-speaking musicians proved to be more successful at identifying and discriminating lexical tones than their non-musician counterparts. This suggests that experience with music pitch processing may facilitate the processing of lexical pitch. Ours is one of an increasing number of studies that have provided counter-evidence to the general theory of dissociation between music and speech processing.

7. Acknowledgements

The authors would like to thank Richard Ashley, Matthew Goldrick, Balaji Mani, and Purav Jesrani for their assistance in this research. In addition, the first author would like to offer a very special thank-you to her niece, Anna Irene, who at the age of 2 years provided the inspiration for this study. This work is supported by Northwestern University.

8. References

[1] Zatorre, R.J., Belin, P., and Penhume, V.B. "Structure and function of auditory cortex: music and speech", *Trends in Cognitive Science*, 6:37-46, 2000.

[2] Fodor, J., *The Modularity of Mind,* MIT Press, Cambridge, 1983.

[3] Bever, T.G., "Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: holistic and analytic", *Annals of the New York Academy of Sciences* 163:251-262, 1975.

[4] Bever, T.G., and Chiarello, R.J., "Cerebral Dominance in Musicians and Nonmusicians", *Science*, 185:537-539.

[5] Kimura, D., "Cerebral dominance and the perception of verbal stimuli", *Canadian Journal of Psychology*, 15:166–171, 1961.

[6] Kimura, D., "Left-right differences in the perception of melodies", *Quarterly Journal of Experimental Psychology,* 16:335–358, 1964.

[7] Curry, F. K.W., "A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks", *Cortex,* 3:343–352, 1967.

[8] Shankweiler, D., and Studdert-Kennedy, M., "Identification of consonants and vowels presented to left and right ears", *Quarterly Journal of Experimental Psychology,* 19:59–63, 1967.

[9] Studdert-Kennedy, M., and Shankweiler, D., "Hemispheric specialization for speech perception", *Journal of the Acoustical Society of America,* 48:579–594, 1970.

[10] Van Lancker, D., and Fromkin, V.A., "Hemispheric specialization for pitch and 'tone': Evidence from Thai", *Journal of Phonetics,* 1:101–109, 1973.

[11] Bryden, M.P., *Laterality: Functional asymmetry in the brain,* Academic Press, New York, 1982.

[12] Peretz, I., "Processing of local and global musical information in unilateral brain-damaged patients", *Brain,* 113:1185-1205, 1990.

[13] Ip, K.F., and Hoosain, R., "Dichotic listening of Chinese and English words", *Psychologia,* 36:140–143, 1993.

[14] Wang, Y., Jongman, A., and Sereno, J.A., "Dichotic Perception of Mandarin Tones by Chinese and American Listeners", *Brain and Language* 78:332-348, 2001.

[15] Patel, A.D., "Language, music, syntax, and the brain", *Nature Neuroscience*, 6:674-681, 2003.

[16] Boersma, P., and Weeknik, D., "Praat: Doing phonetics by computer", http://www.fon.hum.uva.nl/praat/, v. 4.3.04, 2005.

[17] Blicher, D. L., Diehl, R., and Cohen, L.B., "Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement", *Journal of Phonetics*, 18:37-49, 1990.

[18] Schneider, W., Eschman, A., and Zuccolotto, A., "E-Prime Reference Guide", Psychology Software Tools, Pittsburgh, PA, 2002.

[19] Gottfried, T.L., and Riester, D., "Relation of pitch glide perception and Mandarin tone identification", *Journal of the Acoustical Society of America,* 108:2604, 2000.

[20] Gottfried, T.L., Staby, A.M., and Ziemer, C.J., "Musical experience and Mandarin tone discrimination and imitation", *Journal of the Acoustical Society of America,* 115:2545, 2004.

[21] Peretz, I., "Auditory atonalia for melodies", *Cognitive Neuropsychology,* 10:21-56, 1993.

[22] Wong, P.C.M. & Perrachione, T.K. "Training English-speaking adults to use Mandarin tones lexically," *Proceedings of Interspeech 2005 – Eurospeech -- 9th European Conference on Speech Communication and Technology*, submitted for publication.