Towards the role of working memory in pitch processing in language and music

Leigh VanHandel[1], Jennie Wakefield, and Wendy K. Wilkins[2]

Response paper to "Music, language and modularity in action" by Isabelle Peretz, in *Language and music as cognitive systems,* edited by P. Rebuschat, Martin Rohrmeier, John Hawkins and Ian Cross for Oxford University Press.

[1] Assistant Professor of Music Theory and Cognitive Science, Michigan State University; East Lansing, MI; address correspondence to lvh@msu.edu.

[2] Provost and Vice President for Academic Affairs, University of North Texas; Denton, TX

In her article in this volume, as well as elsewhere, Peretz discusses the concept of domain specificity and its relationship to music and language processing. Along with Patel (this volume), Peretz's concern is whether, and to what extent, music and language processing rely on shared mechanisms within the brain. One of these shared resources is working memory, which has received increased attention in cognition literature.

This response describes a recent behavioral experiment carried out at Michigan State University which indicates that working memory plays an important and multi-faceted role in the processing of pitch in language and music, and illustrates the need to further explore the relationship between the mechanisms of working memory and pitch processing in language and music.

The Tone Project

The Tone Project was an investigation of linguistic and musical tone processing that began with a consideration of congenital amusia as described by Peretz and colleagues (Ayotte *et al*, 2002 and elsewhere). We were interested in the general observation that individuals who have difficulties with distinguishing pitch contours in music are not generally thought to have any sort of problem with pitch distinctions in language.

There have been suggestions that this music-language difference could be due to the fact that linguistic pitch contours, such as those involved in sentence prosody, are less subtle than those relevant for music (Peretz and Hyde 2003). We were interested, therefore, in constructing a study in which musical and linguistic pitch contours could be closely matched, in order to carefully examine the similarities and differences in these two closely related cognitive domains.

Further, because it is known that pitch, generally referred to as tone, can be a distinctive contrastive feature in so-called tone languages, we investigated tone perception in music and language and in two distinctively different languages, English (a stress-using language) and Mandarin (a contour tone language).

The study

We recruited 48 native speakers of English and 48 speakers of Mandarin Chinese. All reported normal hearing, normal language skills and acquisition, and were not music majors or professional musicians, although some subjects did self-report having some musical training. Instruction and practice for each task was provided in the native language of each subject, and order was random within tasks and counterbalanced across tasks. Data presentation and response collection was done via e-Prime.

Our experimental paradigm was influenced by the nature of the same-different melody tasks developed initially for the testing of acquired amusia, and then used (in the form of the Montreal Battery of Evaluation of Amusia, the MBEA) in testing for congenital amusia. Our experiment consisted of four same/different tasks using two-note dyads, longer melodies, individual two-syllable words, and longer sentences.

The linguistic stimuli consisted of speech samples, recorded by native speakers of the language, of two-syllable words or multi-word sentences. Target words were recorded one syllable at a time and then combined into words.

For the Chinese stimuli, they were manipulated to remove any redundant phonetic information that might normally accompany pitch contour changes. (For detailed information on the manipulations of the Chinese stimuli, see Meng *et al,* 2005.)

For English stimuli, the manipulation and synthesis involved decomposing stress into its component parts (pitch contour, syllable duration, vowel quality alterations, and loudness), allowing the pitch contour to vary but keeping all other features of stress identical across the two syllables.

The two-syllable word pitch contours were manipulated to match the pitch intervals in the music dyad stimuli. The sentence-level contours were maintained so as to be as natural as possible, and only the pitch of the target word was altered. Subjects made same/different judgments with respect to stimulus pairs that were either identical or that were minimally different and involved only the single localized pitch distinction.

The music tasks consisted of same/different pairs of melodic two-note dyads and relatively short melodies. The pitch intervals used for the dyads were matched to pitch contours used for the word pairs. For the melody task, some melodies were borrowed directly from the MBEA, while others were modified versions composed specifically for this study so that we had a sufficient number of stimuli without repeating any. The melodies were selected or written to be studied in comparison to the sentence-level pitch contours.

Studying the role of working memory

When doing a same/different task such as those in the MBEA and our study, there are two types of errors that can be made. One error is if a subject calls two stimuli different when they are actually the same; this is a *false alarm*. The other possible error is if a subject calls two stimuli the same when they are actually different; this is a *miss*.

Our subjective experience of the MBEA led us to doubt that the mistake of a *false alarm* on “same” stimuli was equivalent to the mistake of a *miss* on “different” stimuli. We believed that a mistake of a *miss* is not necessarily an inability to recognize a pitch distinction; instead, it would seem to be a problem with remembering the first stimuli while listening to the second. We thought it might be revealing to consider subjects’ working memory abilities in addition to their pitch discrimination capabilities; to test these, we selected one verbal working memory task and one spatial working memory task, not knowing which, if either, might be relevant.

To study verbal working memory, subjects were administered a reading span task via computer that required them to respond via paper and pencil. Subjects were presented with a letter flashed on the screen, followed by a distractor sentence which they had to judge (via computer input) for meaning and sense. This pattern would then repeat. After 2-5 iterations of the pattern, subjects were asked to write down the letters they had seen in order. The letters used were Roman alphabet letters for both language groups, but the sentences were translated to Chinese for the Chinese subjects.

The spatial working memory task, a symmetry span test, was also administered via computer and also required subjects to respond using paper and pencil. Subjects were instructed to recall both the direction and length of an arrow originating from the center of concentric circles. The length of the arrow was always short (to the inner circle) or long (to the outer circle), and the direction of the arrow varied throughout the full 360 degrees in 45-degree increments.

Subjects were shown this figure on the computer, immediately followed by a distractor task, which was also a spatial task involving identifying whether letters had been shown in a rotated or mirror-imaged and rotated version. Subjects answered the distractor task via computer, and after between two to five iterations of this sequence, subjects were asked to draw the direction and length of the arrows they had seen on their paper response sheet.

Results

The study resulted in 47 sets of usable data for the English subjects and 46 for the Chinese.

Figure 1 about here

For Chinese speakers, performance scores for accuracy were higher on language than on music tasks for both the words/dyads and sentences/melodies tasks, with no correlation between performance scores on language and music tasks.

In contrast, native English speakers had a higher level of accuracy on music tasks relative to language for both the words/dyads and the sentences/melodies tasks, and demonstrated a strong positive correlation between performance on music tasks and language tasks. This suggests that English speakers process musical and linguistic pitch information in similar fashion for both short and long stimuli, in contrast to the Chinese speakers, whose processing strategies diverge for music and language tone processing.

The finding that Chinese subjects performed much better on the language tasks than the English subjects was exactly what we expected; given that Chinese is a tone language, and that we were studying pitch discriminations, the language task was trivial for the Chinese speakers. However, the Chinese subjects' sensitivity to pitch in language did not seem to translate to music; the Chinese subjects did not perform as well as English subjects on the dyads task, though they did perform better on the melodic task than the English subjects. This may indicate that the Chinese subjects process pitch information differently for short stimuli than for longer stimuli.

As mentioned, subjects were also given tests of working memory designed to determine the effect of specific types of working memory on both the language and music tasks. We hypothesized that if subjects were relying on working memory to complete the tasks, their performance scores would positively correlate with one or more working memory performance scores, and that partialing out the effects of working memory might change the intertask relationships.

Towards the role of working memory

In the original characterization of congenital amusia (Ayotte *et al.* 2002), those who performed at three standard deviations below the control group mean on at least two of the six subsets in the MBEA were categorized as having congenital amusia.

We were also interested in comparing high and low performers, but we realized that using only a subject's total score was not giving a complete picture of performance. For example, it would be possible for a subject to correctly identify 100% of the *same* stimuli but only correctly identify 60% of the *different* stimuli, or having a 40% *miss* rate. This would result in a subject's overall score of 80% correct not reflecting the wide disparity in performance on the two types of tasks. Indeed, we disaggregated our data and found that high performance on the *same* stimuli was not a guarantee of high performance on the *different* stimuli, and that using only the total score hid some people who had no trouble with the *same* stimuli but who struggled to correctly identify the *differents*.

Rather than deciding *a priori* how to group our subjects, we used a k-means based cluster analysis, using as input both the same and different scores for dyads, melodies, words, and sentences, as well as the scores on the working memory tasks. A k-means cluster analysis looks for patterns among the subjects, and groups together those who performed similarly.

Three natural groups for each native language resulted from the cluster analysis. Even though the data for determining the clusters had included the results for the same/different tasks for all music and language tasks and the working memory scores, the results of the cluster analysis isolated one group of subjects in each language who had overall performed poorly on the *different* music tasks.

Figure 2 about here

This chart shows each group's mean performance and standard deviation for the *different* tasks; in each case with an asterisk, the group's performance on the different stimuli for that task was significantly lower (at $p < .05$) than the other groups' mean performance. The English subjects in Group 3 performed significantly lower than the other two groups on both the music and the language tasks, while the Chinese subjects in Group 3 performed significantly lower on the music tasks alone, reinforcing the interpretation that Chinese subjects may have been processing the music and language tasks differently. Even for the Chinese subjects who struggled with the music *different* tasks, the language tasks were trivial; while the Chinese Group 3 did perform slightly less well on the *different* sentences task, it was not significantly different from the other groups.

These results made us want to look specifically at the profiles of people who were scoring well or poorly on pitch in the context of the longer melodic task -- that is, the on the task most like the MBEA -- but for obvious reasons we didn't want to rely only on the total percent correct scores.

For each subject we tallied their responses to the melodic stimuli in four ways: total % correct, % correct of the *same* stimuli, % correct of the *different* stimuli, and a d' score, a single accuracy score that reflects a balance between *same* and *different* scores. We submitted these four measures to another k-means cluster analysis, this time basing the clusters only on subjects' scores on melody tasks.

Figure 3 about here

This k-means cluster analysis provided a high group and a low group for both English and Chinese; the chart shows the mean and standard deviations for the two groups resulting from the these four melody scores. For each task, the mean for the low group is significantly different (at $p < .001$) than the mean for the high group of the same language.

Distinguishing the low and high groups for melody did not create natural classes with respect to our working memory tasks; the individual working memory scores are not significantly different in the low and high groups overall based on this grouping.

However, we found that when we partialed out the subjects' scores for the spatial working memory test regarding arrow length, there was a distinction between the low and high groups in both languages in terms of a music-related d' score and a separate working memory score, namely arrow direction.

Figure 4 about here

For the English subjects who struggled with the melody task, the melody d' score shows a significant positive correlation with the spatial working memory task focusing on arrow direction, while for their Chinese counterparts, we see a significant positive correlation between the music dyads task d' score and arrow direction on the spatial working memory task. While these relationships are different, they both illustrate a relationship between a music-related task and a spatial working memory task, in this case direction.

With these same low and high groups, we also looked at the effect of controlling for performance on the spatial direction task (rather than the spatial length task). For the low-performing Chinese subjects, the d' score for the sentence task is highly significantly positively correlated with both the verbal working memory task and arrow length in the spatial working memory task.

As discussed previously, the results shown in Figure 2 implied that Chinese subjects who were prone to *miss* errors were differentially processing the short stimuli and the long stimuli. Figure 4 illustrates that the subjects in the low-performing Chinese group are, in fact, using different working memory strategies for the two significant interactions. For the short condition, the dyads task showed a significant correlation with scores on the direction element of the spatial task, while the words task did not have enough variance to be able to determine any relationship. (Recall that even the low performing Chinese subjects scored at or near 100% on the words task, as shown in Figure 2.) For the long condition, the low-performing Chinese group demonstrated a significant correlation between the sentences task and both the verbal memory task and the length element of the spatial task.

Discussion

Our results indicate that research on tone or pitch discrimination are probably not applicable cross-linguistically, or at least not for stress and tone languages. We believe that it is likely that speakers of stress-based languages like English process pitch in language and music in a more similar fashion than do speakers of tone languages.

Our results also indicate that specific types of working memory seems to be a factor in pitch discrimination tasks for both language and music for at least a subset of subjects. Therefore, research on any element of pitch discrimination must take into account subjects' working memory abilities not just for music recall, but for other types of working memory. While we don't yet know what the full role of working memory is in pitch discrimination, we do know that partialing out working memory scores changes the intertask relationships, and we know that it interacts differently with pitch discrimination in low and high scorers. Further study into the role of various types of working memory on pitch discrimination is clearly needed.

We have also shown the importance of disaggregating subjects' data; *false alarms* appear to have different origins than *misses,* and this phenomenon needs to be studied more carefully with respect to language and music. In addition, the disaggregation allows a cluster analysis to group subjects naturally to highlight intertask relationships that might otherwise have been hidden.

Much of Peretz's discussion in this volume focuses on the relationship between a specific type of music (singing) and language action (i.e., the actual act of speech); it remains to be seen what the role is of working memory in a study of the relationship between speech production and vocal music production.

Acknowledgements

This work was supported by a Michigan State University Incubator Grant. We thank Dennie Hoopingarner, David McFarlane, Deborah Moriarty, Brad Rakerd, Frederick Tims, Kyle Grove, Matthew Husband, Yuanliang Meng, and Denise Travis for their invaluable assistance in this project.

Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain,* Vol. 125, No. 2, 238-251.

Meng, Y., Rakerd, B., and Wilkins, W. (2005). Disaggregating co-occuring phonetic features: Examples from speech synthesis of Mandarin Chinese. Paper presented at the Michigan Linguistics Society, East Lansing MI, October 2005.

Patel, A., Foxton, J., and Griffiths, T. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* 59(3): 310-13.

Peretz, I. and Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Science* Vol. 7, No. 8, pp. 362-367.

Schellenberg, E.G., and Peretz, I. (2008) Music, language and cognition: unresolved issues. *Trends in Cognitive Sciences* Vol. 12 No. 2, 45-46.