Differences in Brain Responses to Vowels and Musical Intervals

Elika Bergelson[1], Michael Shvartsman[2], and William Idsardi[3]

[1]University of Pennsylvania, [2]University of Michigan, [3]University of Maryland

BACKGROUND AND AIMS

Music and language share many properties: both are complex acoustic signals that unfold over time, and consist of small units grouped into larger, complex units [1]. We investigated whether the brain processes two-formant vowels and frequency matched two-note musical intervals in the same way. While much work has examined the interaction of these complex elements in music cognition [2] and vowel processing [3], neural investigation and comparison of the basic units of music and language is still in its infancy (though see [4-8]). Using magnetoencephalography (MEG), we compared the mismatch-response (MMN/MMF, an early, automatic, pre-attentive difference detector [5] generally occurring approximately 200ms post-onset) to music intervals and vowels composed of matched frequencies. Our broad aim was to determine whether two intervals instantiated as music or language would affect how the similarity between them is reflected in the MMN. That is, would the MMN reflect a psychoacoustic grouping, or would the behavioural and neural signals compute similarity differently, perhaps due to the pre-attentive nature of the MMN [9]. Summarily, our results show that within 250 milliseconds of hearing a complex auditory stimulus, the brain rapidly calculates the relationship between this stimulus and a preceding one, with the electrophysiological response of what makes stimuli 'similar', just as the behavioural one, differing significantly depending on whether the sounds are classified as music or language.

METHOD

Participants. 55 adult volunteers participated in the study, 18 were excluded for various reasons (non-compliance, non-native English speakers, lack of identifiable response to the pre-test, equipment failure) leaving 37 usable participants (18 female; mean age 23.5 years). All participants gave written informed consent. Each session lasted for 60–90 min. The involvement of human participants in the reported experiment was approved by the IRB of the University of Maryland (College Park, Maryland, USA). Subjects either received course participation credit or were paid for their participation.

Stimuli. Auditory stimuli were of 3 types: pure sinusoids (s), piano tones (p), and two-formant vowels (v). Each stimulus featured two simultaneous pitches that had identical first "formants" (i.e. strong resonances) (C4/523 Hz) and second formants creating three ratios: a Major 7th (B4/987 Hz, 11 semi-tones apart), an octave (C5/1046 Hz, 12 semi-tones apart) and a two-octave span (C6/2092Hz, 24 semi-tones apart). All stimuli were digitized at 44.1 kHz, lasted 250ms, and had average amplitudes of 75 dB. The vowels were low-pass filtered at 3000Hz to remove the higher formants. Phenomenologically, the sinewave and piano intervals of M7, P8, and P16 all sound like musical intervals. The vowel intervals of M7 and P8 sound like mid-back vowels (variants of [ʌ], as in "cut") whereas the vowel interval of P16 is perceived as the front vowel [ɛ], as in "get".

Procedure. Participants lay in supine position in a magnetically shielded room while MEG recordings were acquired. Earphones delivered the auditory stimuli binaurally. Auditory stimuli were presented in four blocks of 7.5 min each. Each participant heard two different intervals in an oddball paradigm: a series of standards with interleaved infrequent deviants. The standard to deviant ratio was 6:1, with the number of standards in a row varied from 4-7. Which dyad served as the deviant switched across the two blocks of the experiment. The ISI varied randomly between 500-1000ms. Participants heard 771 standards and 129 deviants in each dyad order. Participants lay passively and could elect to watch a video (with no audio). Each participant heard only two dyads, and the type (vowel, sine, piano) was held constant within subjects (i.e. a subject heard either v8 and v7, or s8 and s16, etc.) Thus, this experiment has a between-subjects design for the contrasts of interests, although each subject does contribute data for both the right and left hemispheres, and in both directions.

Processing. The signal was sampled from DC to 500Hz with an online 200Hz lowpass filter and 60Hz notch filter. Data were noise-reduced offline. The root mean square average of channels selected from an auditory pretest was used for analyses of the MMN. The standard responses were subtracted from the deviant responses to get the mismatch response. We then measured the average amplitude of the mismatch response and the peak latency in the window from 150-250ms. The obtained amplitude and latency responses were then analyzed statistically. We expect to find that the amplitude of the mismatch response should track the perceptual distance or dissonance within the pairs. That is, vowels should show larger mismatch responses to the comparison between octave intervals and two octave intervals (which sound like different vowels) whereas the piano and sinewave tones should show larger mismatch responses to the comparison between octave intervals and sevenths (because the seventh is a more dissonant contrast to the octave than the two octave interval is).

RESULTS

Statistical analyses were conducted using JMP 8 (SAS Inc.) and R (http://www.r-project.org) on each hemisphere separately because the cortical sources in the two hemispheres must be generated separately. Linear mixed effects models were fit for the mismatch field amplitude, with subject as a random effect. The fixed effects were type (piano, sine, vowel), compared intervals (octave-seventh, octave-two octave) and direction (octave standard, octave deviant), and the interactions of the fixed effects. In both hemispheres the main findings matched the experimental predictions: the only statistically significant effect was the interaction of type and interval (LH: $F(2,30) = 5.70$, $p < 0.008$, RH: $F(2,30) = 4.83$, $p < 0.015$). All other main effects and interactions were non-significant ($p > 0.05$). The lack of an effect for direction suggests that the mismatch responses in these conditions were symmetric (that is, a dissonant interval is just as unexpected in a series of consonant intervals as the reverse). Planned comparisons (sinewave and piano versus vowels) were significant for the octave-seventh contrast in both hemispheres (LH: $F(1,28) = 5.89$, $p < 0.022$, RH: $F(1,29) = 5.67$, $p < 0.024$). The octave-two octave contrast was significant in the left hemisphere ($F(1,30) = 4.79$, $p < 0.037$) but not in the right hemisphere ($F(1,30) = 0.97$, $p > 0.3$), due to the strong response for the sinewave in both contrasts in the right hemisphere. Figure 1 shows the mean mismatch amplitudes for the type and interval interactions in the left and right hemispheres. Linear mixed effects models with the same design were also fit for mismatch field latency. No main effects or interactions were found in the left hemisphere (all $p > 0.05$). In the right hemisphere a main effect was found for type ($F(2, 29) = 10.32$, $p < 0.0004$). All other main effects and interactions were non-significant ($p > 0.05$). Since we had no experimental hypotheses regarding latency, post-hoc tests (Tukey-Kramer Honestly Significant Differences, $\alpha = 0.05$) were run, and show that the vowel latency was shorter than the others, but that there was no difference between the piano and sinewave latencies.

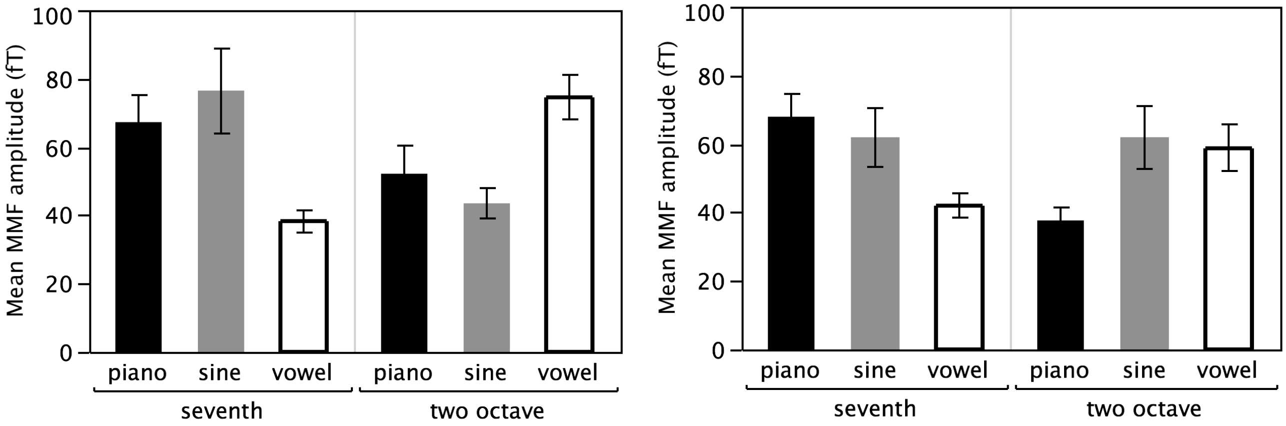

Figure 1: Mean mismatch field amplitudes in femtoTesla for left hemisphere (left) and right hemisphere (right) for each sound type (piano (black), sinewave (grey), vowel (white)) and interval contrast (octave-seventh, octave-two octave).

CONCLUSIONS

As expected, we found a top-down influence of cognitive domain on the low-level auditory processing as indexed by the mismatch response. Musical sounds (in particular piano tones) elicited larger responses to the octave-seventh contrast, in line with the greater dissonance of that pairing. Conversely, vowels elicited larger responses to the octave-two octave contrast, in line with the perceived vowel difference. Thus, the low-level processing of spectral information is influenced by the expectations induced by the cognitive domain of the sounds, suggesting that different perceptual distance metrics influence early spectral auditory processing.

REFERENCES

[1] Patel AD. Music, language, and the brain. Oxford: Oxford University Press; 2008.
[2] Krumhansl CL. Cognitive foundation of musical pitch. Oxford: Oxford University Press; 1990.
[3]Delattre PC, Liberman AM, Cooper FS, Gerstman LJ. An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. Word 1952; 8:195–210.
[4] Kazanina N, Phillips C, Idsardi W. The influence of meaning on the perception of speech sounds. Proc Natl Acad Sci U S A 2006; 103:11381–11386.
[5] Naatanen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 2007; 118:2544–2590.
[6] Tervaniemi M, Kujala A, Alho K, Virtanen J, Ilmoniemi RJ, Naatanen R. Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magnetoencephalographic (MEG) study. Neuroimage 1999; 9:330–336.
[7] Trainor LJ, McDonald KL, Alain C. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. J Cogn Neurosci 2002; 14:430–442.
[8] Zatorre RJ, Belin P, Penhune V. Structure and function of auditory cortex: music and speech. Trends Cogn Sci 2002; 6:37–46.
[9] Jacobsen, T. Mismatch negativity to frequency changes: no evidence from human event-related brain potentials for categorical speech processing of complex tones resembling vowel formant structure. Neurosci Letters 2004; 362: 204–208.

TOPIC AREAS

Music & Language, Music & Neuroscience, Pitch & Tonal Perception