Aspects of Multisensory Perception: The Integration of Visual and Auditory Information in Musical Experiences

By: DONALD A. HODGES, W. DAVID HAIRSTON, AND JONATHAN H. BURDETTE

Hodges, D., Burdette, J. & Hairston, D. Aspects of multisensory perception: The integration of visual and auditory information processing in musical experiences. In G. Avanzini, L. Lopez, S. Koelsch, & M. Majno (Eds.), 175-185. *The Neurosciences and Music II: From Perception to Performance, Annals of the New York Academy of Sciences*, Vol. 1060.

Abstract:
One of the requirements for being a successful musical conductor is to be able to locate sounds instantaneously in time and space. Because this requires the integration of auditory and visual information, the purpose of this study was to examine multisensory processing in conductors and a matched set of control subjects. Subjects participated in a series of behavioral tasks, including pitch discrimination, temporal-order judgment (TOJ), and target localization. Additionally, fMRI scans were done on a subset of subjects who performed a multisensory TOJ task. Analyses of behavioral data indicate that, in the auditory realm, conductors were more accurate in both pitch discrimination and TOJs as well as in locating targets In space. Furthermore, these same subjects also demonstrated a benefit from the combination of auditory and visual information that was not observed in control subjects when locating visual targets. Finally, neural substrates in BA 37, 39/40 were identified as potential areas underlying the conductors' superior multisensory TOJs. Data collection and analyses are ongoing and will lead to an improved understanding of multisensory integration in a complex, musical behavior.
KEYWORDS: multisensory processing; auditory-visual discrimination; temporal-order judgments

Article:
INTRODUCTION
Music making is a prime example of multisensory processing in a complex form of human behavior. At a minimum, information from the ears, eyes, and motor systems must he fully integrated for successful performances. While a considerable amount has been learned in recent years about how the brain organizes musical behaviors, l .2 little attention has been paid to how different systems integrate into a coherent whole. The purpose of this study was to examine multisensory processing in a select group of musicians.

Successful conductors have developed a myriad of skills, including reading a musical score (musical notation indicating the precise notes to be performed by each member of the ensemble), expressing musical information in precise physical gestures (e.g., tempo and dynamics), and retaining the idealized version of the sounds in auditory memory while monitoring the actual sounds produced in real time. Of particular interest is how visual information (e.g., from the score and from the players) is integrated with auditory information (both real and imagined). As just one example, consider that the conductor must be very adept at not only identifying errors, but also locating the errant sound in precise time and space. Thus, it is likely that experienced conductors have developed specialized skills at sound localization, instantly identifying exactly "who" played "what" wrong note. What are the neural substrates behind the visual and auditory processing required for such behaviors? Do experienced conductors possess enhanced neural processing for such tasks?

Multisensory integration is a natural brain function, as input from different sensory modalities is integrated into a coherent perceptual gestalt;[3] each sensory domain can receive input from other senses, and there are transitional multisensory zones between modality-specific cortical domains.[4] This type of integration between senses has been shown to enhance perception in a variety of realms, including speeding up responses,[5],[6] increasing localization accuracy or the detectability of a stimulus,[7-9] and even enhancing perceptions of temporal-order judgments (TOJs).[10,11] However, it must be noted that enhancing effects are typically prominent only when the target stimuli are at or near a perceptual threshold or difficult to locate initially; this effect has become known as "inverse effectiveness," [12] and suggests that the amount of cross-modal benefit observed is directly tied to the relative efficacy of each of the sensory channels involved.

While neuromusical research in general has increased significantly, few studies have investigated conductors specifically. In a study of error detection in conductors,[13,14] PET scans indicated different brain activations for the identification of melodic, harmonic, or rhythmic errors. In another study, event-related brain potentials and behavioral data suggested that conductors were superior to pianists and controls in focusing preattentively and attentively on auditory localization tasks.[15] Although it is clear that a conductor's perceptual gestalt is based on multinodal sensory input, many aspects of multisensory processing in such a complex, dynamic form of behavior remain poorly understood. Thus, it was the purpose of this experiment to investigate aspects of multisensory perception in the integration of visual and auditory information processing in conductors when compared to controls. Reported here are preliminary results from studies currently in progress.

SUBJECTS

Subjects for this experiment consisted of ten conductors and ten musically untrained controls. Conductors were between the ages of 30 and 40 (mean = 34.4, ± 4) and had from 6-18 (mean = 10.35, ± 3.9) years' experience as a middle or high school band or orchestra director. All were right handed, as determined by self-report. Seven of the subjects were male; three were female. These musicians had an average of 19.5 years of education (± 2.3). Control subjects were similar in age (mean = 33.4, ± 4.4), handedness, gender (6 males, 4 females), and general educational background (19 years, ± 3.7). However, they lacked formal music instruction, having minimal or no formal musical training.

METHODS FOR BEHAVIORAL TASKS

Subjects completed a series of behavioral tasks involving unisensory and multi- sensory processing. Some of these tasks were subsequently replicated during fMRI scanning.

Pitch Discrimination

Subjects reported which of two tones occurred first ("high first" versus "low first"), using two buttons on a response box. The order of presentation was always random. The two tones were played with a 500 ms pause between them, and they always differed in pitch. One tone was always 440 Hz and the other was always higher, but the specific frequency varied according to an adaptive staircase procedure, starting at 457 Hz (roughly two thirds the "tonal" distance between A and A#) and adjusting in "distance," and hence difficulty, until subjects were performing at roughly 71% accuracy; eventually the pitches were so close together that subjects could barely distinguish which of the two was the higher one. The end measurement was a frequency value that represented their threshold discrimination above the base tone (e.g., 442.8 Hz). Three such staircase procedures occurred simultaneously, and on each trial one was chosen at random, with the average of all three used as the final measure per subject.

A second task was identical, except that the base frequency was 500 Hz, and comparison tones started at 520 Hz. This was done to determine whether conductors were better with a well-known note (A) versus a tone not represented in the Western musical scale.

Temporal-Order Judgment

In the third task, auditory TOJ, the two tones did not change (440 and 660 Hz, A and E), but the onset time between them changed via an adaptive staircase; that is, they were presented closer together in time until they appeared to occur simultaneously. A determination was made of the time needed to discriminate the two pitches (e.g., 26 ms).

For the fourth task, visual TOJ, subjects determined which circle (of two) presented on a computer screen (200 Hz vertical refresh rate) was presented first—the top circle or the bottom circle. As in previous tasks, the stimulus-onset asynchronies (SOAs) varied according to an adaptive staircase to determine the threshold for visual discrimination.

Temporal-Order Judgment with Multisensory Cues

This TOJ task (visual circles with monotone beeps) was similar to the visual- alone TOJ task, in that the subjects determined which visual circle was presented first, but in this "multisensory" version, they were also presented with nonspatial **10** ms clicks with the presentation of the circles. The first beep was congruent with the first circle, but the second beep could be congruent with the second circle or delayed (50-350 ms). The delay between the auditory beeps was altered in an attempt to determine a window for multisensory integration. Measurements indicated how much change in accuracy occurred in the various sound-added conditions verses no sound, as well as the average time to respond.

Target Localization

Subjects were asked to locate visual (LED), auditory (broadband noise burst), or spatially congruent visual-auditory stimuli across a 180 arc. These stimuli were presented in a dark chamber, and the subjects were asked to locate either a briefly presented (50 ms) light, a sound, or both light and sound using a laser-pointing yoke device. Targets were presented 10, 20, 30, or 40 degrees to the left or right and were randomly interleaved across trials. Responses were measured with regard to both speed and accuracy.

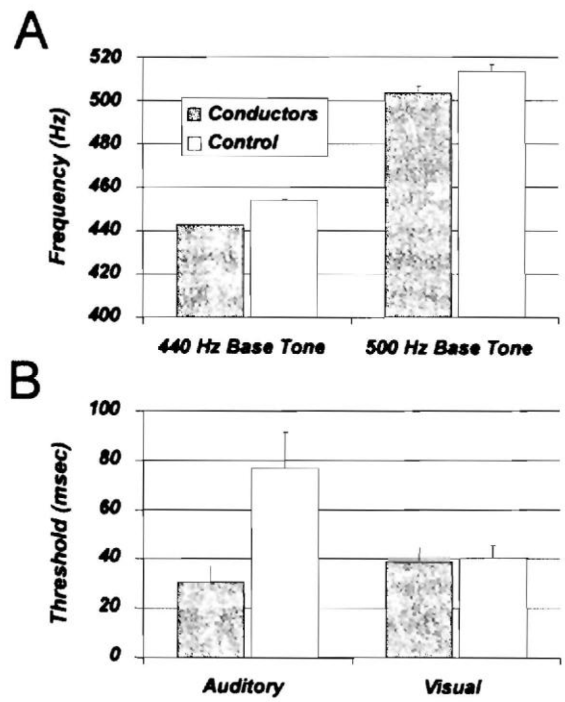

FIGURE 1. (**A**) Results for pitch discrimination tasks. (**B**) Results for auditory and visual temporal-order judgments. Error bars represent SEM for the group.

RESULTS FOR BEHAVIORAL TASKS

Pitch Discrimination

On average, conductors tested thus far are significantly more acute in their ability to discriminate between different pitches than musically untrained individuals. FIGURE. 1 A shows the average pitch that could be discriminated above two different baseline tones (440 and 500 Hz) for each of the two groups (error bars represent SEM). Because this trend was significant for both a baseline of 440 Hz *(418)* = 3.66, $P < .05$). A4 on the equal-tempered musical scale, and 500 Hz *(018)* = 3.24, $P < .05$), which is no particular note but would lay roughly between B4 and C5, it is not likely that this difference is selective to specific notes of the musical scale. As a matter of perspective, while control subjects require a tone to be roughly 2.95% of the base to discriminate it, conductors require a difference of only 0.67%.

Temporal Discrimination

Subjects also performed auditory and visual TOJs in order to ascertain their ability to discern between sounds over time. With regard to temporal discrimination, an interesting dichotomy appears, shown in FIGURE 1B. Specifically, in addition to discriminating between pitches, conductors are significantly more acute in the temporal domain with sounds as well. Control subjects require an average of 76.7 milliseconds between onsets to discriminate the two tones (440 and 660 Hz) at a threshold level, while conductors require only 33.7 ms in order to perform at the same level (418) = 2.87, $P < .05$). In contrast, both groups required approximately the same amount of time when the task involved visual stimuli (40.4 ms and 38.9 ms, respectively; 418) = 0.20, $P > .05$), suggesting that these enhancements are limited to the auditory domain.

Temporal Discrimination with Multisensory Cues

For each subject, the threshold SOA determined from the above visual TOI was used to fix the SOA to a single value within a set of visual TOJs that also included a task-irrelevant sound. Previous studies 1[0,11] have shown that the inclusion of this sound with a slight delay improves the accuracy with which subjects can perform the visual TOJ task. For both groups, this was indeed the case—subjects were significantly more accurate when the second sound was delayed by 100 ms than with either the sound synchronous (conductors: $t(9) = 4.32$, $P < .05$; control: $t(9) = 1.96$, $P = .05$) or no sound at all (conductors: $t(9) = 2.62$, $P < .05$; control: $t(9) = 9.93$, $P < .05$). Additionally, this effect showed a decline with larger delays, such that when the cross-modal SOA was as much as 350 ms, no benefit was observed (conductors: $t(9) = 1.0$, $P > .05$; control: $t(9) = 0.65$, $P > .05$); this is consistent with previous reports. FIGURE 2 shows the average time required to respond for each of the groups across all conditions.

Interestingly, response times are decreased (i.e., faster decisions) for both groups simply by including the sound synchronous with the onset of the two visual stimuli, even though it does not provide any task-relevant information; note that this does not quite reach statistical significance, although a trend is clear (conductors: $t(9) = 2.15$, $P < .016$; control: $t(9) = 1.95$, $P = .083$). Additionally, response times are fastest with

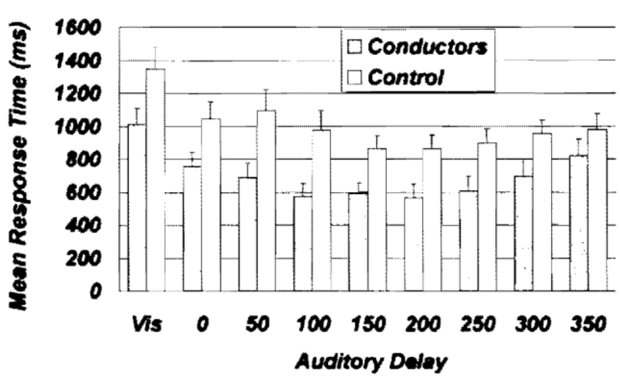

FIGURE 2. Results for visual TOJ when a task-irrelevant sound is added.

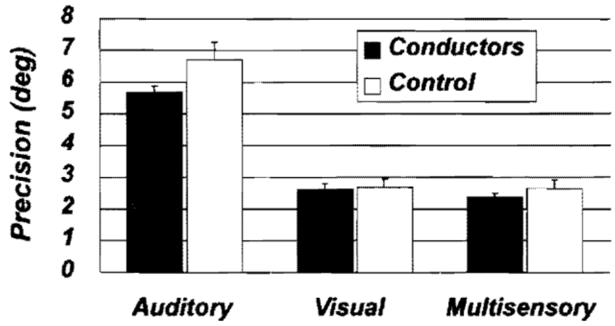

FIGURE 3. Results for localization tasks.

cross-modal SOAs of 150-200 ms between the latter two stimuli, and, as with accuracy, this effect weakens with an increased delay. Notice also that conductors are consistently faster than control subjects across all conditions tested ($F(1, 18) = 5.37$, $P<.05$).

Target Localization

After completion of all threshold testing, we assessed subjects' abilities to locate visual, auditory, and combined visual-auditory stimuli. FIGURE 3 shows the average localization precision for both groups for each target type.

In this case, "precision" is assessed as the average width, in degrees of angle, of the distribution of responses (standard deviation). Hence, low values (e.g., a small number of degrees) represent higher precision, whereas larger values denote less precision. First notice the pattern for control subjects. While their ability to locate *the* auditory target is relatively poor, it is significantly better when the target is visual in nature. However, the addition of an auditory stimulus to this visual target does not significantly improve this ability ($t(9) = 0.29$, $P >$ .05). By contrast, conductors show a different trend. Primarily, their auditory localization ability is noticeably better than that of control subjects (while this does not quite reach statistical significance in a between-groups analysis ($t(18) = 1.8$, $P = .09$), this is likely due to the low number of subjects examined thus far). Additionally, unlike their control counterparts, conductors receive a significant benefit from the additional auditory signal, such that their localization ability is significantly better for multisensory than visual-only targets ($t(9) = 2.65$, $P.< .05$), an effect that is likely due to their enhanced auditory performance. Finally, as with multisensory-mediated TOJs, conductors consistently responded faster than control subjects, showing significantly decreased response times with visual ($t(18) = 2.92$, $P < .05$), auditory ($t(18) = 3.08$, $P < .05$), and multi- sensory ($t(18) =$ 2.99, $P < .05$) stimuli.

fMRI METHODS

We have begun an fMRI study in the same individuals on whom the behavioral tasks have been performed. We are first attempting to image differences in the patterns of activation between the conductors and nonconductors performing the multisensory TOJ task described previously. Specifically, we want to determine which underlying cortical networks are responsible for the speeding of response time in the conductors. Presented here are preliminary data for two conductors and four nonconductors.

fMRI Experimental Paradigm

During the fMRI session, subjects performed the identical multisensory TOJ task using an event-related paradigm. Stimuli were presented through MR-compatible goggles and headphones (<www.mrivideo.com>). Just as in the behavioral study discussed above, in each condition the subject had to state which of two circles appeared first. Two different visual SOAs were used for all conditions: one matched the individual threshold derived during the TOJ testing above, while the other was a constant value for all subjects (50 ms). Additionally, auditory delays were constrained to include only 100 and 300 ms. In addition to two visual-only control conditions, two auditory-only conditions were included, consisting of delays matching the two multisensory conditions but during which subjects did not respond. Together, each of these eight conditions (four multisensory, two visual, two auditory) were presented in a random order 14 times during each of five sessions, for a total of 70 stimulus "events" per condition.

fMRI Image Acquisition and Analysis
All imaging experiments were performed on a GE echo-speed Horizon LX MR scanner with a birdcage head coil (GE Medical Systems, Milwaukee, WI). Whole- brain activation was assessed by examining blood oxygenation level-dependent (BOLD) changes[16],[17] by measuring changes in the T2*-relaxation rate caused by the changes in blood oxygenation that accompany cortical activation.[18,19] Functional imaging was performed in the axial plane using multislice gradient-echo echo-planar imaging (TR = 2500, TE = 40 ms) with a field of view of 24 cm (frequency) x 15 cm (phase), and an acquisition matrix of 64 40 (28 slices, 5 mm thickness, no skip). Statistical parametric maps (SPMs) were generated using SPM99 [20,21] from the Wellcome Department of Cognitive Neurology (London, UK) implemented in Matlab (The Mathworks Inc., Sherbom MA, USA), with an LDL interface. Anatomic regions were defined using an anatomic MRI atlas[22] normalized to the same MNI-SPM template. SPMs were generated by means of the general linear model within SPM99. The data were analyzed using a fixed effects model for each group (conductors and nonconductors) and thresholded at $P <$ .001, corrected for multiple comparisons at $P < .05$.

fMRI RESULTS
When analyzing the four multisensory conditions together, the conductors and nonconductors show, not unexpectedly, activity in the visual and motor cortex. However, the conductors showed greater activity in the visual cortex, especially in the higher-order visual cortices, and extending into the occipitotemporal regions in Brodmann's area (BA) 37 (FIG. 4). We probed the data using a region-of-interest (ROI) analysis constrained to known multisensory/heteromodal brain regions (BA37, 39/40), and the differences between the activation patterns for 1he conductors and nonconductors were more apparent (FIG. 5). Specifically, the conductors showed increased activity in bilateral occipitotemporal cortices in BA37. While the nonconductors did show activity in the higher parietal regions, they did not show increased activity in BA37 areas.

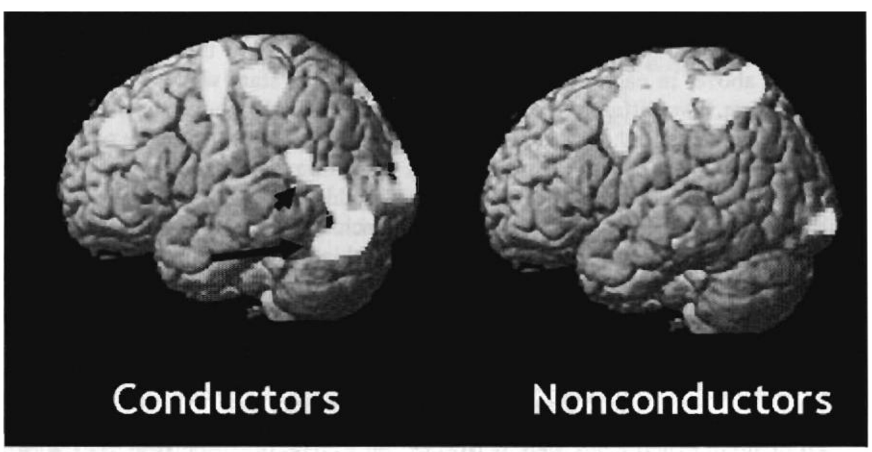

FIGURE 4. Patterns of significant cortical activation for conductors (*left*) and controls (*right*) on multisensory TOJs. Note the increased occipitotemporal activation (*black arrow*) and superior temporal sulcus activation (*arrow head*) in the conductors. Peak activity for the conductors was in the left occipital lobe (Montreal Neurological Institute (MNI) coordinates −8, −108, 10; peak t score = 9.61) and left occipitotemporal region (−20, −76, −15; peak t = 8.18). The peak activity for the controls was in the left parietal lobe (−32, −60, 60; peak t score = 7.36) and right occipital lobe (20, −80, −15; peak t = 6.65).

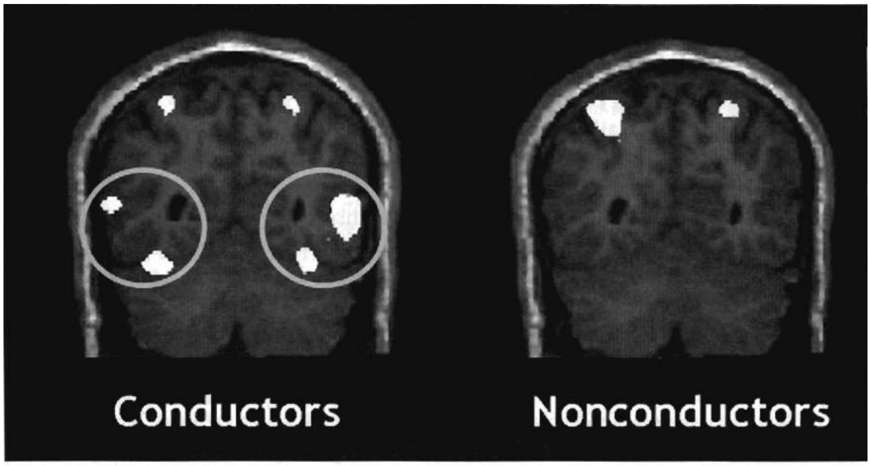

FIGURE 5. Increased activations in bilateral occipitotemporal cortices (*gray circles*) in BA37 for conductors (*left*) but not for controls (*right*). Peak activity for the conductors in the right BA37 was at MNI 56, −60, 10, with peak $t = 7.74$; and in the left BA37 was at −40, −64, −20, with peak $t = 6.22$. Peak activity for the controls was in the left parietal lobe (−32, −48, 45; peak $t = 6.82$).

CONCLUSIONS

Musical behaviors are primary examples of multisensory processing. Among musicians, conductors have a particular need for heightened spatial localization skills. It was the purpose of this experiment to examine multisensory processing in this select group of musicians. In the preliminary data gathered so far, ten experienced conductors and ten controls with very Limited musical training participated in a series of behavioral tasks. Two conductors and four nonconductors subsequently underwent fMRI scans while performing the multisensory TOJ task.

Data analyses indicate that conductors have finely tuned auditory processing skills, including more refined pitch discrimination and shorter auditory temporal thresholds. In other words, they are more accurate in making pitch discriminations than control subjects and require less time between two sounds to be able to discriminate which of them occurred first. Furthermore, they demonstrated clearly improved response times to TOJs in multisensory conditions, when compared to visual alone. That is, besides 1he combination of auditory and visual information improving performance on the task, it also enabled them to provide a response more quickly. Although nonconductors also show such a benefit in their performance, the improvement in response time was less clear, and they were, in general, much slower to respond than were individuals with experience in both music and conducting.

The benefits of multisensory integration were also seen for conductors, but not nonconductors, in the target localization task. The integration of auditory and visual information enhanced performance, allowing conductors, once again, to respond faster and more accurately, while control subjects did not show such a benefit. These results fit nicely with the observation that conductors, while on the podium, must instantaneously be able to locate "who made what sound" Despite the typically seen highly accurate localization ability in the visual-alone condition (seen in both the conductors and nonconductors), the conductors still received a benefit in locating ability when the auditory stimulus was added. While such multisensory enhancement of behavior is usually present only with stimuli near threshold,[r- the conductors received a benefit from the added auditory stimulus even with both the auditory and visual stimuli substantially above threshold detection level. Perhaps the conductors have developed (or were born with?) neural ensembles or connections that allow the greater or more frequent integration of auditory and visual information.

In the first stages of the imaging component of this project, neural substrates underlying the conductors' superior multisensory TOJ performances have been identified. BA37, 39/40 are particularly implicated and are known as multisensory convergence areas involved in other behaviors. The cortex in BA37 at the occipitotemporal junction is a known heteromodal area of audiovisual convergence[23] and, interestingly, has been shown to be an important area in the acquisition of reading skills.[24] These early results may show a neural "signature" of multisensory binding and may show the brain network responsible for a more efficient processing of the simultaneously presented visual and auditory stimuli. Scanning additional subjects will determine whether these results are robust. If so, we will have the beginnings of an understanding of the neural mechanisms underlying multisensory processing in conductors,

REFERENCES

1. AVANZINI, G., C. FAIENZA, D. MINCIACCHI, *et al.*, Eds. 2003. The Neurosciences and Music. Vol. 999. The New York Academy of Sciences. New York, NY.
2. ZATORRE, R.J. & I. PERETZ, Ens. 2001, The Biological Foundations of Music. Vol, 930. The New York Academy of Sciences. New York, NY.
3. WALLACE, M., G. ROBERSON, W. HAIRSTON, *et al.* 2004. Unifying multisensory signals across time and space. Exp. Brain Res. 158: 252-258.
4. WALLACE, M., R. RAMACHANDRAN & B. STEIN. 2004. A revised view of sensory cortical parcellation. Proc. Natl. Acad. Sci. USA 101: 2167-2172.
5. HUGHES, H., P. REUTER-LORENZ , G. NOZAWA & R. FENDRICII. 1994. Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. J. Exp. Psychol. Hum. Percept. Perform. 20: 131-153.
6. HARRINGTON, L. & C. PECK. 1998. Spatial disparity affects visual-auditory interactions in human sensorimotor processing. Exp. Brain. Res. 122: 247-252..
7. HAIRSTON, W,, P, LAURIENTI, G. MISHRA, *et al.* 2003. Multisensory enhancement of localization under conditions of induced myopia. Exp. Brain Res. 152: 404-408.
8. LAURIENTI, P., R. KRAFT, J. MALDJIAN, *et al.* 2004. Semantic congruence is a critical factor in multisensory behavioral performance. Exp. Brain Res. 158: 405A4.
9. LOVELACE, C., B. STEIN & M. WALLACE. 2003. An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Res. Cogn. Brain Res. 17: 447-453.
10. MOREIN-ZAMIR, S., S. SOTO-FARACO & A. KINGSTONE. 2003. Auditory capture of vision: examining temporal ventriloquism. Brain Res. Cogn. Brain Res. 17: 154-163.
11. HAIRSTON, W.D., J.H. BURDETTE, D.L. FLOWERS, *et al.* Abnormal temporal processing of cross-modal information in dyslexia. Exp. Brain Res. In press.
12. STEIN, B. & M. MEREDITH. 1993. *In* The Merging of the Senses. M. Gazzaniga, Ed. MIT Press. Cambridge, MA.
13. PARSONS, L., D. HODGES & P. Fox. 1998. Neural basis of the comprehension of musical harmony, melody, and rhythm. Society for Neuroscience Annual Meeting. J. Cogn. Neurosci. Abstracts.
14. PARSONS, L. 2001. Exploring the function neuroanatomy music performance perception, and comprehension. Ann. N. Y. Acad. Sci. 930: 211-230.
15. NAGER, W., C. KOHLMETZ A. AL*TENMULLER, et al. 2003*. The fate sounds in conductor’s brains: an ERP study Brain Res Cogn Brain Res. 17: 81-93.
16. BUCHBINDER, B. & G. COSGROVE. 1998. Cortical activation MR studies in brain disorders. Magn. Reson. Imaging Clin. N. Am. 6: 67-93.
17. TURNER, R. , A. HOWSEMAN, G. REES, et al. 1998. Functional magnetic resonance imaging of the human brain: data acquisition and analysis. Exp. Brain Res. 123: 5-12,
18. OGAWA, S,.LEE, A. KAY. *et al.* 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA 87: 9868-9872.
19. OGAWA, S., R. MENON. D. TANK, *et al.* 1993. Functional brain mapping by blood oxy genation level -dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. Biophys. J. 64: 803-812.
20. FRISTON, K., C. FRITH, R. FRACKOWIAK. *et al.* 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage 2: 166-172.

21. FRISTON, K..C. FRITH, m TURNER, ,,*al.* 1995. Characterizing evoked hemodynamics with fMRI. Neuroimage 2: 157-165.
22. KINKINIS, R., P. GLEASO*N, T. MORIARTY Et al.* 1996. Computer-assisted interactive three-dimensional planning for neurosurgical procedures. Neurosurgery 38: 640- 649; discussion 649-651.
23. BUCHEL C, C. PRICE & K. FRISTON. 1990. A multimodal language region in the ventral visual pathway. Nature 394: 274-277.
24. PUGH, K., W. MENCL, A. JENNER, *et al.* 2001. Neurobiological studies of reading and reading disability. J. Commun. Disord. 34: 479-492.