10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Multisensory integration in percussion performance

Bruno L. Giordano[1], Federico Avanzini[2], Marcelo Wanderley[1], Stephen McAdams[1]

[1] CIRMMT, Schulich School of Music, McGill University, 550 Sherbrooke West, H3A1E3 Montréal, Quebéc, Canada

{bruno.giordano,marcelo.wanderley,smc}@music.mcgill.ca

[2] Department of Information Engineering, University of Padova, Padova, Italy, federico.avanzini@dei.unipd.it

We investigated how auditory and haptic information concerning objects hardness is integrated for the purpose of controlling the velocity with which we strike an object. Our experimental manipulations and data analyses considered a variety of factors that should be integrated in a theory of multisensory perception: expertise of the perceiver; context (unimodal vs. multimodal); inter-modality congruence; inter-participants agreement in sensory weighting; performance.

On each trial, participants struck a virtual object with a constant target velocity and received feedback on correctness. When the performance criterion was reached, feedback was eliminated, the auditory and/or haptic hardness of the struck object were changed, and the effects on subsequent striking velocity and performance were measured. In unimodal trials only the haptic or auditory display was presented. In multisensory trials, the audio-haptic changes could be congruent (e.g., both increased in hardness) or incongruent. We recruited participants with different levels of expertise with the task: percussionists, nonpercussionist musicians and nonmusicians.

For both modalities, striking velocity increased with decreasing hardness, and vice versa. With the vast majority of participants, changes in haptic hardness were perceptually more relevant because they influenced striking velocity to a greater degree than did changes in auditory hardness. The perceptual weighting of auditory information was robust to context variations (unimodal vs. multimodal), independent of expertise, uniform across participants and modulated by audio-haptic congruence. The perceptual weighting of haptic information was modulated by context and expertise, was more varied across participants and was robust to changes in audio-haptic congruence. Performance in velocity tracking was more strongly affected by haptic than by auditory information, was not at its best in a multisensory context and was independent of information congruence.

1 Introduction

Our environment is populated with multimodal objects and events that have a variety of properties: spatial and temporal, but also nonspatial and nontemporal (e.g., transparency). Perception of these properties serves two types of tasks: nonaction tasks (e.g., How far is that object?) and action tasks (e.g., Grasp that object; see [1] and [2] for differences between action and nonaction perception). Despite the ecological variety of object and event properties and of perceptually guided tasks, research on the integration of multisensory information has predominantly focused on the perception of spatial and temporal properties of objects and events in nonaction contexts (e.g., Is this beep on the right or left of this flash?, [3]). As such, the ecological validity of current theories of multisensory integration has not been extensively assessed.

We conducted an empirical study on the influence of auditory and haptic information on the control of the velocity with which a perceiver strikes an object. We measured the extent to which striking velocity and performance in the control of velocity are influenced by auditory and haptic information about the hardness of an object (see [4] for the multisensory perception of surface properties, and [5] for auditory perception of hardness). We assessed the effect of three different factors likely to affect the strength of the perceptual effects of sensory information: context, i.e., whether modality-specific information is presented in isolation or along with information from different modalities [6]; prior experience with the task, i.e., whether participants were already experienced in the control of striking velocity (percussionist musicians) or not (nonpercussionist musicians and nonmusicians; [4]); congruence of multisensory information, i.e., whether in a multisensory context information from one sensory modality is consistent with information from a different sensory modality [7]. Analyses focused on measures of effects within populations of participants and also on measures of interindividual differences.

2 Experiment

Participants were asked to repeatedly strike a simulated audiohaptic object with a velocity that was within a target range. The perceptual relevance of sensory information was measured by the effect of a change in the properties of the simulated object on motor behavior, i.e., on the striking velocity itself, and on performance, i.e., on measures of the ability to keep the striking velocity within the target velocity range.

2.1 Methods

2.1.1 Participants

Forty-two right-handed observers took part in the experiment (26 females, 16, males; mean age = 22.5 yrs, STD = 4). They all had normal hearing, as assessed with a standard audiometric procedure [8, 9]. Fourteen participants were nonmusicians (less than four years of musical training, mean = 0.9 yrs; STD = 1.4); 14 were musicians with no training in percussions(at least four years of musical training, mean = 12.6 yrs; STD = 5.1); 14 were musicians with a minimum of 5 years of training in percussion (mean = 11.6 yrs; STD = 3.7).

2.1.2 Apparatus

Participants were seated inside an IAC double-walled soundproof booth. They held with their right hand the stylus of a Phantom Desktop, a force-feedback device used to present the haptic stimulus. The Phantom Desktop was interfaced with a real-time model for the synthesis of impact sounds [10, 11], and with a software system for the control of the properties of the haptic stimulus [12]. Sound signals (44.1 kHz, 16 bit, peak intensity = 75 dB SPL ca., as measured with a Brüel & Kjær Type 2205 sound-level meter coupled with a Brüel & Kjær Type 4153 artificial ear) were amplified with a Grace Design m904 monitor system, connected to the optical port of the PC workstation used to control the experiment, and presented binaurally through Sennheiser HD280 headphones.

In order to minimize kinesthetic information about striking velocity, the right arm of the participants was strapped to the right arm of the chair so that only wrist movements could be used to displace the stylus of the force-feedback device. No visual information about striking velocity was provided: both the Phantom Desktop and the right arm of the participants were hidden from view.

2.1.3 Stimuli

Sounds were synthesized using a real-time physically inspired model of a square plate with five vibrational modes and with material properties similar to those of hard wood (frequency of the lowest vibrational mode = 100 Hz; [10]). The amplitude and spectral properties of the presented impact sounds were modified based on the velocity with which the stylus of the Phantom Desktop reached a given spatial position. Across trials, we manipulated a mechanical parameter of the impacted sound source and of the simulated haptic object: the force stiffness coefficient K measures the compression of the striking object produced by a given striking force [13], and its value increases with an increase in the stiffness of both of the objects in collision. From the auditory standpoint, higher values of this parameter determine an increase in the perceived hardness of both the striking and struck object [5].

Impact forces were haptically simulated using a linear stiffness model, which generated a reaction force proportional to the normal displacement of the stylus relative to the surface of the virtual plate. An ideally rigid wall should be simulated with the highest possible stiffness. In practice, however, limitations in the haptic device restrict the available range of values for the simulated stiffness, because exceedingly high values can result in unstable interactions. In order to improve the stability of the interaction, we added a dissipative component to the contact force [14]. All the haptic feedback was programmed with the Openhaptics^{TM} Toolkit developed by Sensable.

2.1.4 Procedure

Each trial was divided into three phases. During an initial training phase, participants were instructed to strike the virtual object with a constant target velocity. For each strike, they received feedback about whether the striking velocity was below, above or within a target velocity range (430-570 mm/s). This phase ended after five consecutive correct strikes. During a subsequent adaptation phase, participants were asked to continue striking with the same target velocity in absence of feedback. This phase ended after five strikes, independently of their correctness. At the beginning of a final change phase, the value of the audio and/or haptic K was modified. Participants were instructed to continue striking with the same velocity, and to ignore the changes in the properties of the simulated object. No feedback on performance was given. This phase ended after 20 strikes, independently of whether they were correct or not. Participants were instructed to strike with a tempo of their choice, provided that they kept it constant throughout an entire trial.

We investigated three experimental conditions. In the haptics condition, we manipulated only haptic information. During this condition, participants heard a continuous white masking noise (level = 75 dB SPL ca.). During the auditory condition, we manipulated only acoustical information. When the stylus reached the contact position, the same as during the other conditions, an impact sound was synthesized based on the impact velocity and was played back. No haptic stimulus was presented. During the change phase of both the haptic and auditory conditions, K could assume one of five log-spaced values centered around the baseline value used during the training phase of each trial (haptic K: 93-1860 N/m, baseline = 416 N/m; acoustical K: 1,000-100,000 N/m$^{1.5}$, baseline: 10,000 N/m$^{1.5}$). During the multisensory condition, we manipulated both the haptic and acoustical stimulus. During the change phase of this condition, we combined factorially all the levels of acoustical and haptic K from the auditory and haptic conditions. In the congruent multisensory condition, both the acoustical and haptic K either increased or decreased relative to the baseline level. In the incongruent multisensory condition, the acoustical and haptic K changed in opposite directions (e.g., increase in acoustical K and decrease in haptic K).

In each block of 15 trials, each of the possible levels of the change-phase K for each of the experimental conditions were presented once (five haptic, five auditory and five multisensory trials). Each block of 15

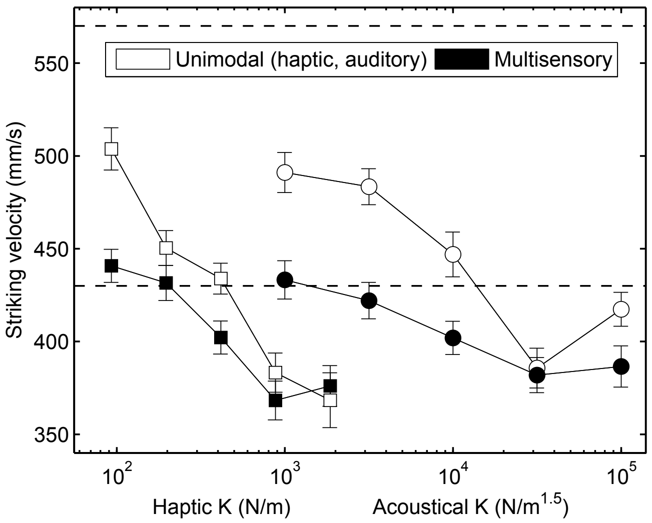

Figure 1: Across-trials average of the change-phase striking velocity for one experimental participant as a function of haptic and acoustical K (square and circles, respectively) in the unimodal and multisensory conditions. Dashed lines bracket the target velocity range. Error bars = ± 1 SE. Note that haptic and acoustical K have different units of measure.

trials was subdivided into five subblocks of three trials each. During each subblock, participants were presented with one trial from each of the three experimental conditions (random order of conditions within each subblock). Each participant completed 20 blocks of 15 trials each for a total of 600 trials. Throughout all the multisensory trials, each of the possible combinations of the change-phase nonbaseline haptic and acoustical K values was presented five times, whereas the baseline value for the haptic and acoustical K was presented 20 times. Each participant completed the experiment in three experimental sessions of 1.5 hours each on different days.

2.2 Results

Figure 1 shows the across-trial average of the change-phase striking velocity for one of the experimental participants in each of the experimental conditions. We carried out two different analyses, modeling the effect of the experimental factors on either the change-phase striking velocity, or on the number of correct strikes in the same phase of a trial. In both cases, data were analyzed with linear mixed-effects models (LMMs; [15, 16]), a powerful extension of the general linear model. In general, LMMs can include both fixed effects, describing the relationship between independent and dependent variables in the entire population, and random effects, quantifying the variability of the effects of the independent variables within units of interest (in this study, the population of participants). More specifically, the LMMs presented in this paper model the variation of the participant-specific effects within the population of participants as a normal probability density function with a specific mean and variance: the average value of the normal distribution corresponds to the fixed effect and estimates the average effect in the population; the variance of the normal distribution corresponds to the random effect, and estimates the extent to which a specific effect differs across individuals. Differently from classical linear analyses such as ANOVA, with LMMs not significant effects are discarded, and a the structure of the covariance matrices (e.g., covariance matrix of the model residuals) is modeled so as to better fit the input data. For all of the models presented in this paper, effects were selected using the top-down strategy described in [16].

2.2.1 Striking velocity

Within a first LMM (Fig. 2, left panel), we modeled striking velocity data from all of the experimental conditions. For each of the participants, we considered the average of the change-phase striking velocity across repeated presentations of the same level of K (haptic and auditory conditions) or repeated presentations of the same combination of haptic and acoustical K (multisensory condition) for a total of 27 data points for each of the 42 participants. We tested for significant fixed effects of experimental condition (auditory, haptic, multisensory), expertise (nonmusician, musician, percussionist), log-transformed haptic and acoustical K, and all the possible interactions between these factors. We also tested for random effects of haptic and acoustical K and expertise-related differences in the random effect of haptic and acoustical K. The value of the log-transformed haptic and acoustical K was standardized prior to being entered into the LMM. For this reason, the absolute value of the estimates of effect of these variables can be interpreted as a measure of effect size. This standardization also makes the estimates of the random effects for the same variables comparable. The data of four participants were not included in the final model because they violated the assumption of normality of the participant-specific residuals ([17], Shapiro-Wilk $W \leq .923$, $p \leq .046$). The final model explained 78% of the variance in the input data and included significant fixed effects of acoustical K and experimental condition, the interaction between acoustical and haptic K, the interaction between experimental condition, on the one hand, and haptic K or expertise, on the other, and the interaction between haptic K and expertise, $F \geq 3.91$, $p \leq .020$, $p > .05$ for the not significant fixed effects. Importantly, the effect of acoustical K was modulated neither by the experimental condition nor by expertise. Overall, higher values of acoustical K thus induced a slower striking velocity, and vice versa. On the contrary, the effect of haptic K was modulated by both the experimental condition and by expertise: as compared to the unimodal haptic condition, in the multisensory condition higher values of haptic K tended to produce faster striking velocities. Also, whereas percussionists struck faster for higher values of haptic K, both nonpercussionist musicians and nonmusicians struck more slowly for higher values of the same variable (see Fig. 2, left panel). The final model also included a significant random effect of haptic K, $\chi^2(1, 2) = 293.9$, $p < .001$[1], whereas the random effect of acoustical K was not significant, $\chi^2(2, 3) = 2.00$, $p = .470$. Based on this result, a

[1]The reference distribution for testing the significance of random effects in a LMM is a mixture of two χ^2 distribution with equal weight (0.5). Hence the two degrees of freedom reported for this test.

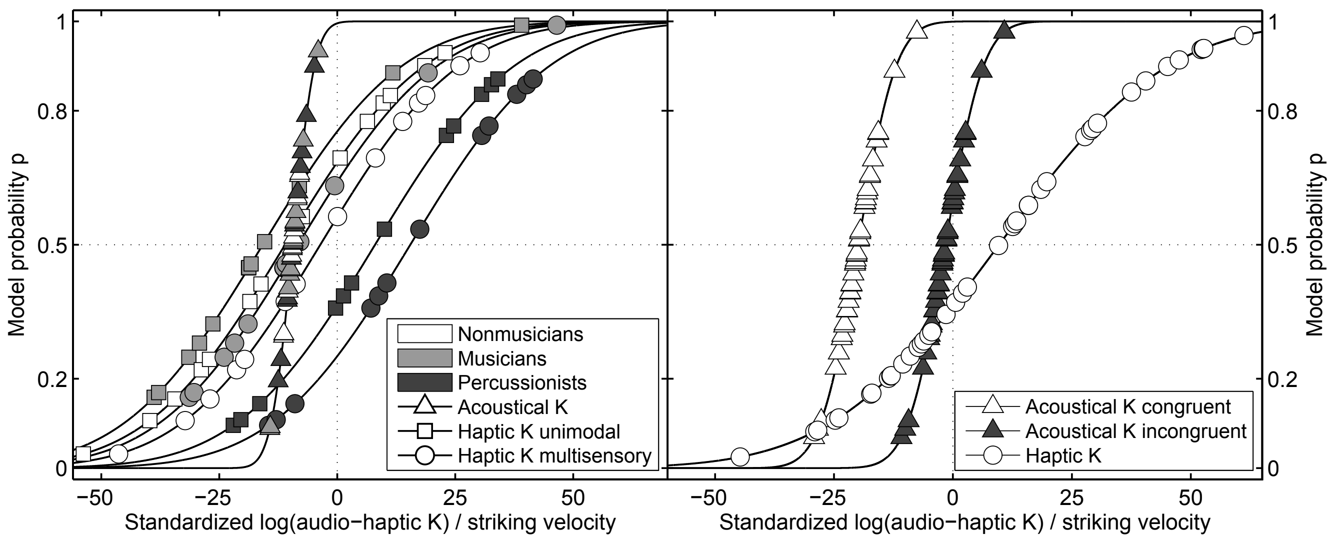

Figure 2: Linear mixed models (LMMs) of striking velocity in all experimental conditions (left panel) and in the multisensory condition (right panel). Both figures show the estimate of the participant-specific effects in the LMMs described in the text. Fixed effect estimates correspond to the x-axis value for which the cumulative normal distributions reach the 0.5 probability; random effect estimates corresponds to the variance, i.e., spread, of the cumulative normal distributions. For negative estimates of the effect of K, striking velocity decreases with increasing K values and vice versa (Fig. 1), whereas for positive estimates of the effect of K, higher K values produce faster striking velocities and vice versa. For the model presented in the left panel the random effect of acoustical K was not significant, i.e., was not statistically different from zero. For illustrative purposes, we show here the nonzero LMM estimate of this parameter.

larger degree of interindividual differences emerged for the effects of haptic K than for those of acoustical K. Finally, no random effect was significantly modulated by expertise, $\chi^2(5) = 3.80$, $p = .579$, indicating the same level of interindividual differences within nonmusician, musician and percussionist groups.

We assessed whether haptic or auditory information influenced striking velocity most strongly in the multisensory condition, i.e., which sensory modality dominated the multisensory context. To this purpose, we analyzed the absolute value of the multisensory-context participant-specific LMM estimates of the effect of haptic and acoustical K within a repeated measures ANOVA. We included sensory modality as within-subjects factor (haptic vs. auditory), and expertise as between-subjects factor. The effect of sensory modality was significant, $F(1, 35) = 28.79$, $p < .001$, $\eta_p^2 = .451$, showing that changes in haptic K affected striking velocity more strongly than changes in acoustical K. Neither the effect of expertise nor that of the interaction of this factor with sensory modality was significant, $F(2, 35) \leq 15.34$, $p \geq .191$, $\eta_p^2 \leq .011$, indicating that the strength of the effect of both acoustical and haptic K was not modulated by expertise.

Within a second LMM (Fig. 2, right panel), we modeled striking velocity data from the multisensory condition. For each participant, we considered the average change-phase striking velocity across the repetitions of each of the 16 possible combinations of the nonbaseline change-phase levels of haptic and acoustical K. Within this model, we tested for a significant fixed effect of expertise, of the log-transformed acoustical and haptic K, of the congruence of audio and haptic K, and of all the possible interactions with these factors. We also tested for a random effect of haptic and acoustical K and for expertise-related differences in the random effects for these factors. As with the previous LMM model, the log-transformed values of haptic and acoustical K were standardized prior to entering them into the LMM model. The residuals for four participants violated the normality assumption, Shapiro-Wilk $W \leq .875$, $p \leq .032$. Even when the data for these participants were removed, the structure of the final model did not change. For this reason, we report here the results for the full dataset. The final model explained 84% of the variance of the input data, and included a fixed effect for haptic and acoustical K, and for expertise and congruence, $F \geq 3.29$, $p \leq .048$. Interestingly, whereas the fixed effect of the interaction between haptic K and congruence was not significant, $F(1, 543) = 0.01$ $p = .937$, that of the interaction between acoustical K and congruence was significant, $F(1, 543) = 19.20$ $p < .001$. As such, the effect of haptic K was independent of whether this parameter was modified in the same direction as acoustical K. On the contrary, the fixed effect of acoustical K was negative when it was changed in the same direction as haptic K, whereas the same parameter had overall no effect on striking velocity when it changed in the direction opposite to that of the haptic K change (see Fig. 2). Also importantly, no significant interaction emerged between congruence and expertise, $F \leq 1.71$, $p \geq .182$, indicating that the congruence of multisensory information had the same effect independently of the level of previous experience with percussion performance. Finally, the selected LMM included a significant random effect of both haptic and acoustical K, $\chi^2 \geq 292.1$, $p < .001$. Consistently with the model for the data from all conditions, a higher degree of interindividual differences emerged for the effect of haptic K than for the effect of acoustical K,

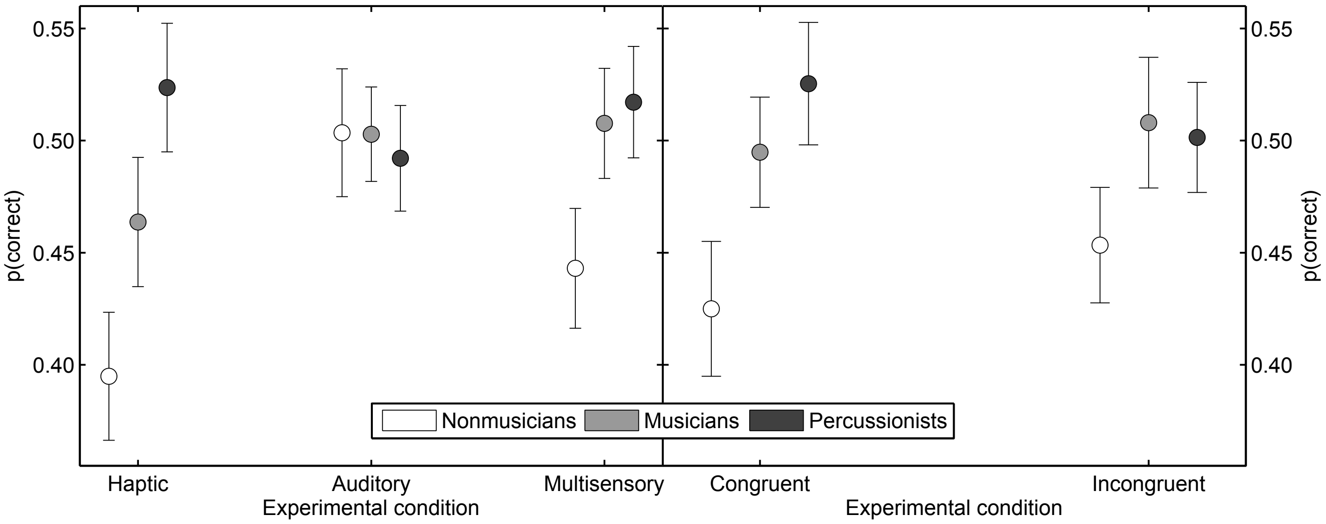

Figure 3: Across-participants average striking performance in all experimental conditions (left panel), and in the multisensory condition (right panel). Error bar = ± 1 SE.

random effect estimate = 771.34 and 38.94, respectively. Also consistently with the model for the data from all conditions, the level of interindividual differences was the same among nonmusicians, musicians and percussionists, $\chi^2(5) = 1.50$, $p = .913$.

2.2.2 Performance

We computed two different LMMs to assess the effect of the experimental manipulations on the measures of striking performance (see Figure 3). Performance was defined as the proportion of change-phase strikes within the target velocity range. For both models, the independent variables haptic and acoustical K were considered as categorical, because their effect on performance were expected to be nonlinear.

The first LMM considered data from all the experimental conditions, and investigated the same effects as the analogue LMM for striking velocity data. Participant-specific data were aggregated across repeated presentations of the same change-phase level of haptic or acoustical K in the unimodal conditions, and across repeated presentations of the same pair of values of haptic and acoustical K in the multisensory condition. The final model explained 70% of the variance of the input data. No participant violated the assumption of normality of the residuals. The final model included significant fixed effects of experimental condition, haptic K and the interaction between expertise and experimental condition, $F \geq 6.03, p < .001$. In particular, a significant pairwise performance difference between expertise-related groups emerged only in the haptic condition, where nonmusicians performed significantly worse than percussionists, $F(1, 168) = 10.73$, Bonferroni-corrected $p = .008$, $F(1, 168) \leq 5.29$, Bonferroni-corrected $p \geq .132$ for the other pairwise contrasts. Consistently with the overall weak effect of acoustical K on striking velocity, the effect of this factor on striking performance was not significant, $F(4, 164) = 1.11$, $p = .353$, $p > .05$ for the other effects not included in the final model.

A final LMM model considered only data from the multisensory-condition trials where the change-phase audio and haptic K differed from the baseline level for the training phase. This model mirrored that used to investigate the effect of multisensory congruence on striking velocity. The final model explained 66% of the variance of the input data. The residuals for three participants violated the normality assumption, Shapiro-Wilk $W \leq .872$, $p \leq .029$. Their data were not considered for further modeling. The final model included a significant effect of acoustical and haptic K, and a significant interaction between expertise and audiohaptic congruence, $F \geq 4.50$, $p \leq 0.012$. We initially investigated the interaction between expertise and audiohaptic congruence by testing for a significant effect of congruence within each of the three expertise-related groups of participants. The effect of congruence was not significant for all three groups of participants, $F(1, 140) \leq 5.04$, Bonferroni-corrected $p \geq .081$. Post-hoc pairwise contrasts between expertise-related groups in the congruent and incongruent conditions revealed instead that in the congruent condition nonmusicians performed significantly worse than percussionists, $F(1, 154) = 7.29$, Bonferroni-corrected $p = .046$, $F(1, 154) \leq 2.46$, Bonferroni-corrected $p \geq .712$ for the other pairwise contrasts.

3 Conclusions

We investigated the integration of auditory and haptic information in the control of the velocity with which an object is struck. For all participants, in the multisensory context haptics and audition emerged as the dominant and secondary modality, respectively.

Information from the least relevant modality, audition: (1) appeared to be processed robustly independently of contextual influences, i.e., had the same behavioral effects in the unimodal and multisensory context; (2) was processed uniformly across participants, and (3) independently of the level of expertise with the task; (4) had an effect on performance that was secondary, at best; (5) was processed differently depending on whether it was congruent with information from the dominant modality or not. The perceptual processing of information from the dominant modality, haptics: (1) was strongly influenced by changes in context, i.e. depended on whether it was presented in a unimodal or multisensory context; (2) largely differed across participants, and (3) was influenced by the level of previous experience with the task; (4) strongly influenced performance levels; (5) was independent of multisensory congruence.

Acknowledgments

This work was supported by Stephen McAdams' Canada Research Chair in Music Perception and Cognition, by a grant to him from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN 312774-05), by a Special Research Opportunity Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) awarded to him and to Marcelo Wanderley. Federico Avanzini was supported by the EU FP6 Network of Excellence "Enactive Interfaces" (IST-1-002114).

The authors wish to thank Paolo Crosato and Stephen Sinclair for their help with the coding of the experimental software, and Mattson Ogg for his help with running experiment participants

References

[1] Aglioti S., DeSouza J. F. X., Goodale M. A., "Size-contrast illusions deceive the eye but not the hand" *Current Biology* 6, 679-685 (1995)

[2] Ellis R. R., Flanagan J. R., Lederman S. J., "The influence of visual illusions on grasp position" *Experimental Brain Research* 2, 109-114 (1999)

[3] Calvert G., Spence C., Stein B., *Handbook of multisensory processes*, MIT Press, Cambridge, MA (2004)

[4] Lederman S. J., Klatzky R. J., "Multisensory texture perception" in Calvert G., Spence C., Stein B. (Eds.), *Handbook of multisensory processes* MIT Press, Cambridge, MA, 107-122 (2004)

[5] Giordano, B. L., Rocchesso, D., McAdams, S., "Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability" *Journal of Experimental Psychology: Human Perception and Performance* (in press)

[6] Ernst M. O., Banks M. S., "Humans integrate visual and haptic information in a statistically optimal fashion" *Nature* 415, 429-433 (2002)

[7] Jacobs R. A., "What determines visual cue reliability?" *Trends in Cognitive Sciences* 6, 345-350 (2002)

[8] International Organization for Standardization, *Acoustics - Reference zero for the calibration of audiometric equipment - Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (ISO 389-8)* Geneva, Switzerland (2004)

[9] Martin F. N., Champlin C. A., "Reconsidering the limits of normal hearing" *Journal of the American Academy of Audiology* 11, 64-66 (2000)

[10] Avanzini F., Rath M., and Rocchesso D., "Low-level sound models: resonators, interactions, surface textures" in Rocchesso D., Fontana F., (Eds.) *The Sounding Object* Mondo Estremo, Firenze, Italy, pp. 119-148 (2003)

[11] Avanzini F., Crosato P., "Integrating physically based sound models in a multimodal rendering architecture" *Computer Animation and Virtual Worlds* 17, 411-419 (2006)

[12] Itkowitz B., Handley J., Zhu W., "The OpenHaptics Toolkit: A Library for Adding 3D Touch Navigation and Haptics to Graphics Applications" *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems: World Haptics Conference* 590-591, Pisa, Italy (2005)

[13] Chaigne A., Doutaut V., "Numerical simulations of xylophones. I. Time–domain modeling of the vibrating bars" *The Journal of the Acoustical Society of America* 101, 539-557 (1997)

[14] Colgate J.E., Brown J.M., "Factors affecting the Z-Width of a haptic display" *Proceedings IEEE International Conference on Robotics and Automation* 3205-3210, San Diego, CA (1994)

[15] Verbeke G., Molenberghs G., *Linear mixed models for longitudinal data*, Springer-Verlag, New York, NY (2000)

[16] West B. T., Welch K. B., Galecki A. T., *Linear mixed models: a practical guide using statistical software*, Chapman & Hall/CRC Press, Boca Raton, FL (2006)

[17] Pinheiro J. C., Bates D. M., *Mixed–effects models in S and S-Plus*, Springer-Verlag, Berlin, Germany (2000)