The Dominance of Haptics Over Audition in Stabilizing Wrist Kinematics During Striking Movements

by

Yinan Cao

Music Technology Area, Department of Music Research
Schulich School of Music
McGill University, Montreal
April 2015

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Arts

© Copyright 2015
by
Yinan Cao

To scientific truth (at least its maximum-likelihood approximation), Anyan, and music.

Abstract

Skilled interactions with sounding objects such as drumming rely on resolving the uncertainty in the feedback from acoustical and tactual signals radiated from the objects being set into vibration. The uncertainty may arise from misestimation of the objects' mechanical properties, such as their stiffness. How the integrated processing of object-generated auditory and haptic information feeds back into the fine-tuning of sound-generating actions remains unexplored. Participants held a stylus and learned to stabilize the downward wrist velocity while striking repeatedly a virtual sounding object whose surface stiffness was under computer control. No visual input about the movements or object was provided. Unbeknownst to participants, the sensory feedback was manipulated by perturbing the acoustic and haptic stiffness of the object either in a congruent or incongruent manner. The compensatory changes in striking velocity were measured as a motor effect of the sensory perturbation. We quantified the contributions of each sensory modality to the control of wrist kinematics by characterizing a potential asymmetry of *congruency effects* across audition and haptics. These effects were operationally defined as the difference in motor effects between incongruent and congruent trials. We found a pronounced dominance of haptics over audition in the stabilization of wrist kinematics. The intersensory weighting required for motor compensation was in accord with a statistically optimal prediction based on modality-specific motor noise. Percussion expertise facilitated motor accuracy (lower motor noise) in all experimental contexts. The action-based multisensory process(es), however, appear to rely upon a minimum-uncertainty mechanism that optimizes the control signal for reliable striking kinematics independently of musical expertise.

Résumé

Des interactions expertes avec des objets sonores, tel que le jeu des instruments à percussion, dépendent de la résolution d'incertitudes concernant la rétroaction provenant de signaux acoustiques et tactiles rayonnés par les objets mis en vibration. Cette incertitude pourrait découler d'une mauvaise estimation des propriétés mécaniques de l'objet, telle que sa rigidité. La façon dont le traitement intégré des informations auditives et haptiques générées par l'objet agit sur l'ajustement des actions qui produisent du son reste inexplorée. Les participants tenaient un stylet et ont appris à stabiliser la vitesse du poignée en frappant de façon répétée sur un objet sonore virtuel dont la rigidité de surface était contrôlée par un ordinateur. A l'insu des participants, la rétroaction sensorielle a été manipulée en perturbant la rigidité acoustique et haptique de l'objet, de manière congrue ou incongrue. Les changements compensatoires dans la vitesse de frappe ont été mesurés en tant qu'effet moteur de la perturbation sensorielle. Nous avons quantifié les contributions de chaque modalité sensorielle au contrôle de la cinématique du poignée en caractérisant l'asymétrie potentielle des *effets de congruence* entre l'audition et le toucher. Ces effets sont opérationnalisés en tant que différence des effets moteurs entre les essais congrus et incongrus. Nous avons trouvé une dominance marquée du toucher par rapport à l'audition dans la stabilisation de la cinématique du poignée. La pondération inter-sensorielle nécessaire pour la compensation motrice était conforme avec une prédiction statistique optimale fondée sur le bruit moteur spécifique à chaque modalité. L'expertise en percussion a facilité la précision motrice (diminution du bruit moteur) dans toutes les conditions expérimentales. Pourtant, le processus multi-sensoriel lié à l'action semble dépendre d'un mécanisme à incertitude minimale qui optimise le signal de contrôle pour la fiabilité de la cinématique de frappe indépendamment de l'expertise musicale.

Acknowledgments

My greatest thanks go to Professor Stephen McAdams and Dr. Bruno Giordano for consistently supporting me (scientifically, financially, and emotionally) throughout my last three years as their student. I was previously trained as an engineer. “A sort of romantic notion about pure logic,” as Professor Brenda Milner once expressed, guided me towards experimental psychology. I sincerely doubt that I would have been able to complete this work without the trust and encouragement of these two supervisors. Their scientific insights, clarity of judgement and thought, and creative ideas/new ways of solving old problems are a continual source of delight and surprise. I doubt that I could have been able to complete this work without their substantial support during the processes of my funding applications. These include an NSERC-Create Auditory Cognitive Neuroscience graduate fellowship, and a CIRMMT Inter-Centre Research Exchange Funding. Professor McAdams also supported me with his NSERC grants. In *The Importance of Stupidity in Scientific Research*, Dr. Martin Schwartz said, “Science makes me feel stupid [...].” No doubt, doing scientific research most of the time feels like “looking for a black cat in a dark room,” all by yourself. What these supervisors have taught me is that while trying to get comfortable with being ‘stupid’, it is important to enjoy the stupidity and find the true meaning of crossing the ocean to understand how the musical brain works.

My thanks also go to Professor Fabrice Marandola (Schulich School of Music, Percussion Area) for providing insights into the interpretation of experimental results; to Professor Marcelo Wanderley for lending me the haptic equipment, and to Professor Gary Scavone for his generosity with the sharing of his (expensive!) data-acquisition card for this research; as well as to Stephen Sinclair and Marcello Giordano (Input Devices and Music Interaction Lab) and Hossein Mansour (Computational Acoustic Modeling Lab) for their helpful instructions on the measurement of motion acceleration.

I've also been fortunate to have supporters outside of McGill. My gratitude extends to Professor Pascal Belin for his generous support on my academic exchange in the Voice Neurocognition Laboratory (University of Glasgow, Scotland) in which I could work with Dr. Bruno Giordano. I would also like to extend a special thanks to Professor Christoph Kayser who provided inspiring comments on an earlier version of this work.

I would like to extend a special thanks to my MPCL-mates present and past including Chelsea Douglas and David Sears in our 'mixed-model panel'; Béatrice Copps, Cecilia Taher, Chris Wood, Dominique Beauregard Cazabon, Eddy Kazazis, Emily Wilson, Hauke Egermann, Kai Siedenburg, Jason Noble, Meghan Goodchild, SongHui Chon, Sven-Amin Lembke, and others for always lending an open ear to presentations, reading manuscripts, and providing advice of all kinds; thanks to Bennett 'the King' Smith, one of the most humorous lab managers and computer geeks in the world—I could not have asked for better colleagues.

Finally, thanks to my family living in the *Celestial Empire* (Ti*ā*nCh*á*o) and friends in Montreal, Dr. Pan Liu and Dr. Xiaoming Jiang (the Neuropragmatics and Emotion Lab, McGill), and Dr. Lixing Zhou (the Quantitative Methods Lab, McGill), for endless love and encouragement.

Author contributions

The work based on the current thesis has been submitted for publication in a manuscript of the same title coauthored by Yinan Cao (printed last name: Tsao), Bruno L. Giordano (Marie Curie Research Fellow, Institute of Neuroscience and Psychology, University of Glasgow), Federico Avanzini (Assistant Professor, Department of Information Engineering, University of Padova) and Stephen McAdams (Professor, Department of Music Research, McGill University).

Among the above coauthors, S. McAdams functioned as the thesis supervisor as well as the director of the laboratory in which the main parts of the research were conducted. His contribution concerns all stages of the work, from overseeing the conception and design of experimental paradigms, discussing the analysis approaches, interpreting the results, and revising the manuscript, to financing the research with regard to technical facilities and the remuneration of participants. The work was supported by a Canadian Natural Sciences and Engineering Research Council grant (NSERC RGPIN 312774–10) and by a Special Research Opportunity Grant from NSERC awarded to S. McAdams, as well as his Canada Research Chair. My contribution as the first author involves performing all reported statistical and acoustical analyses, redirecting the interpretation of the results into new avenues in terms of the computational modeling of motor statistics and motor-based sensory weighting, as well as authoring all parts of this thesis, composing the majority of the manuscript, incorporating revisions suggested by the coauthors, and preparing the paper for submission. B.L. Giordano served as an external thesis advisor. His contribution includes motivating the earlier study conception and design, gathering the data, guiding the analysis approaches and data interpretation, editing the manuscript, as well as supervising me as visiting postgraduate in the Voice Neurocognition Lab (Glasgow) in which I completed portions of the reported linear mixed-effects modeling. F. Avanzini gave technical support and conceptual advice to the experiment and data acquisition, contributed to the

interpretation of the results, and provided critical revision to the manuscript. The work was also financed by the EU FP6 Network of Excellence “Enactive Interfaces” (IST–1–002114) granted to F. Avanzini.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA	Analysis of variance
CDF	Cumulative distribution function
ERBr	Equivalent Rectangular Bandwidth rate
ED	Effective Duration
LMM	Linear mixed-effects model
ML	Maximum likelihood
PC	Principal component
PCA	Principal component analysis
p.s.	pseudo–sone
SCG	Spectral Center of Gravity
SEM	Standard error of mean
SD	Standard deviation
SPL	Sound pressure level

Chapter 1

Introduction

1.1 Sensory prediction error as learning signal for motor adaptation

A crucial problem faced by the sensorimotor control system is inevitable delays in receiving afferent action-related sensory feedback that will be used to guide subsequent motor planning (delays on the order of 100 ms, Saijo, Murakami, Nishida & Gomi, 2005; Franklin & Wolpert, 2008, 2011). For understanding how such a problem can be resolved in the human central nervous system, there is a broad consensus on the computationally efficient nature of a forward (internal) model, which predicts sensory effects of the motor commands driving the coordination of muscle coactivation to achieve the goal(s) of given motor acts, e.g., precise timing or velocity control of a tennis serve (Miall & Wolpert, 1996; Wolpert, Miall & Kawato, 1998). In order to accurately estimate relevant features of movements, such as the time-varying positions or kinematics of the moving limbs or body, the causal relations between the actions and their sensory consequences must be stored in the nervous system to allow for prediction (Wolpert & Kawato, 1998). In essence, the predictive mechanism relies on an "efference copy" of the motor commands that could be further processed to anticipate the action outcomes before the

actual arrival of the delayed sensory feedback (Angel, 1976; Guthrie, Porter & Sparks, 1983).

Evidence favoring the existence of the forward models in state estimation has been found empirically at behavioral levels (Wolpert, Ghahramani & Jordan, 1995). Other evidence from neuroimaging (Blakemore, Frith & Wolpert, 2001; Kawato, Kuroda, Imamizu, Nakano, Miyauchi & Yoshioka, 2003) or patient studies (Serrien & Wiesendanger, 1999; Nowak, Hermsdörfer, Rost, Timmann & Topka, 2004; Tseng, Diedrichsen, Krakauer, Shadmehr & Bastian, 2007) has consistently emphasized the role of the cerebellar circuits in computing forward-model estimates of action-related sensory feedback during predictive sensorimotor control. For people with a structurally and functionally intact cerebellum, their control of grip force is an adaptive process involving efficient anticipation of movement-induced mechanical contact events during object manipulations or explorations. When a grasped object is lifted from a support (such as a table), the mechanical event such as the moment of object lift-off is encoded via tactile signals from the fingertips. Lift-off will occur earlier (or later) than predicted, if an object is lighter (or heavier) than expected, and the resulting sensory mismatch will trigger predictively a decrease (or an increase) in both load (vertical lifting) force and grip force (Johansson & Westling, 1988). Another appropriate example is the preparatory adjustment of grip force when sensing object stiffness with a rigid stylus. The grip force applied to the stylus is observed to be efficiently coordinated with the resistive force generated by the tapped or struck object so as to achieve a trade-off between preventing the stylus from slipping from the grasp and alleviating fatigue caused by an excessive grip force (LaMotte, 2000). This evidence uncovers the reliance of sensorimotor control upon an internal model of the manipulated/explored objects, i.e., one associated with memory-based sensory representations of object properties (Flanagan & Wing, 1997). The predictive control of grip force (or other motor parameters in other behavioral tasks, e.g., stability-related postural force for bimanual movements, limb dynamics for goal-directed reaching, etc.) was instead significantly impaired in patients with cerebellar degeneration (Nowak

et al., 2004; Diedrichsen, Verstynen, Lehman & Ivry, 2005; Smith & Shadmehr, 2005).

Because the nervous system appears to incorporate the feedforward (internal) process(es) that monitor the sensory consequences of the motor commands, how does it improve the motor commands for subsequent actions when receiving unexpected sensory feedback? Sensory prediction errors, which are a discrepancy between predicted and observed sensory representations of performing a motor act, could be used as a potential learning signal for adaptive motor responses (Miall & Wolpert, 1996; Shadmehr, Smith & Krakauer, 2010). Prediction error conceptually corresponds to the "exafferent signal" (Sperry, 1950; Von Holst, 1954), which is the consequence of subtracting the sensory prediction for the planned movements from the actual afferent feedback transmitted through the sensory apparatus. This subtraction process is possibly associated with error-encoded neural activities within the cerebellum (Schlerf, Ivry & Diedrichsen, 2012). Consider a drumming task in which the goal of repeated strokes with a hand-held drumstick on a percussion surface (e.g., drumhead) is to achieve desired percussive sounds with constant loudness and timbre (e.g., brightness). An inexperienced beginner sometimes gets a tinny tone and an abrupt collision force if the stroke accidentally lands close to the edge of the drumhead while originally aiming at its more elastic central area. The errors (the undesired sound and vibration) could therefore be signaled through the unexpected sensory events. The encoding of the impact-related prediction errors would trigger either a return of future movements back to the desired trajectory leading to the correct end point, or an adjustment of the muscle strength applied to subsequent strokes.

The concept of monitoring how sensory perturbation is counteracted by compensatory motor responses has been widely accepted for studying the error-driven motor learning and adaptation within a broad range of sensorimotor activities. Traditionally many previous studies have investigated how sensory prediction error drives motor adaptation by imposing two distinct types of disturbance effects on the sensorimotor control loop. The first type alters the

transformation from motor to concurrent sensory representation, i.e., it perturbs the sensory feedback online as a movement progresses. Exemplar paradigms include not only force-field (Shadmehr & Mussa-Ivaldi, 1994) or visuomotor (Tseng et al., 2007) adaptation of goal-directed arm movements, but real-time perturbations of acoustical feedback and/or somatosensory feedback (path of jaw motion) during vocal (speech and singing) production (Villacorta, Perkell & Guenther, 2007; Tremblay, Shiller & Ostry, 2003; Lametti, Nasir & Ostry, 2012; Jones & Keough, 2008; Keough & Jones, 2009; Zarate & Zatorre, 2008).

Alternatively, sensory prediction errors can be imposed by introducing surreptitious fluctuations in the properties of the object(s) with which the mid-movement limb(s) or a hand-held intermediary object, such as a rigid probe, is (are) about to interact (Castiello, Giordano, Begliomini, Ansuini & Grassi, 2010; Sedda, Monaco, Bottini & Goodale, 2011). Consider a scene in which you reach out to lift from a table a bottle that was supposed to be empty but in fact is not. The potential misestimation for the weight of the lifted bottle might at first cause a slip of the hand from the bottle. Such a prediction error could be detected and resolved so quickly that an increase in the grip force would occur even within the ongoing grasp. From an ecologically valid perspective, prediction errors arising from misinterpretation of objects' properties may naturally come from multiple sensory modalities. By perturbing an object's intrinsic property that can be perceived through different sensory channels, the changes in interactive kinematics (i.e., motor effects of perturbed sensation) can be used to explore the issue of motor-based multisensory processes. For example, the size of an object can be coded tactually, visually, or even auditorily such as by means of the impact sound generated when an object is placed on a table (Grassi, 2005). Using natural impact sounds, Sedda et al. (2011) found that presenting a sound stimulus that carries object-size-related acoustical cues immediately before grasping movements, could affect the grip aperture between the thumb and index finger. There seems to exist a close correspondence between the perceptually estimated size of the object

and the motor commands planned for an appropriate scaling of the grasp aperture. Given this correspondence, Sedda et al. (2011) found a nearly balanced weighting of vision and audition by quantifying the extent to which the noise in motor statistics (standard deviations in maximal grip aperture) matches a maximum-likelihood (ML) prediction in which sensory weighting is determined by sensory reliability, i.e., the inverse of variance due to sensory noise. Note, however, that another crucial hypothesis of the ML principle (see section 1.3 for details) had not been tested in their study (Sedda et al., 2011): the reliability of the multisensory estimate of a physical property must be superior to any of the unisensory estimates.

The significant role played by the auditory processing of altered object-related information of a to-be-grasped target in modulating reaching kinematics has also been unveiled in a previous study by Castiello et al. (2010). The motor task employed in their study was reaching to grasp objects that were visually covered with various materials differing in brittleness, such as aluminum, paper, string, and wool. The sensory perturbation effects were introduced as a crossmodal mismatch between the visually and auditorily coded material properties of the target objects. On each trial, a pre-recorded contact sound generated with the bare fingers was either congruent or incongruent with the visual material, e.g., a "string" sound combined with a visual target covered with aluminum. The sound was presented along with the visual target before or after a reaching movement. The effects of crossmodal facilitation and interference on contact-point variability were revealed for the congruent and incongruent trials, respectively, indicating the existence of salient audio-visual interactions during the control of reach-to-grasp movements. However the relative contributions of each sensory modality were not quantified by Castiello et al. (2010) in that particular study. Nor did they consider the potential role of haptics (active touch) in influencing auditory and/or visual processing of the grasped object's material, e.g., by reinforcing the processing of a congruent feature in another modality (Ernst, Banks & Bülthoff, 2000).

1.2 Uncertainty arising from multiple sources of prediction errors

Consider a goblet that accidentally breaks when clinked with another one during a toast. This event can be a result of fragility (bad quality) of the glass or an excessive force applied by the hand holding it. That is, the exact cause of a prediction error might be ambiguous. Both the uncertainty arising from the fluctuating self-generated movements due to motor variability and the uncertainty arising from the unstable environment can act to disturb a sensorimotor control process. Such a causal ambiguity of the prediction errors in fact serves as the pivot of the reactive feedback mechanisms in most motor adaptation studies using the perturbation paradigms that either perturb the movements or the objects upon which the movements operate. If a prediction error arises, it is necessary to determine whether to assign the error to the dynamics of our own limb(s) or the effects of environmental disturbances resulting from unstable object properties in the extrapersonal space. Uncertainty arising from multiple sources of prediction errors poses a problem of "credit assignment" in the nervous system in which the weight assigned to each source of error is determined in order to adapt control appropriately (Berniker & Körding, 2008; Wei & Körding, 2009). An understanding of the process(es) by which motor adaptation operates has been postulated by Berniker & Körding (2008)'s optimal modeling of error source estimation using Bayesian statistics. Using reaching tasks, Wei & Körding (2009) provided a probabilistic schema explaining the empirical evidence that the larger the sensory mismatch created by an excessively high degree of visuomotor distortion, the less it was likely to be causally relevant to movement production and, consequently, to be adapted to in a linear manner. Causal ambiguity of error source thus appears to be more pronounced when it is difficult to distinguish between intrinsic (limb) and extrinsic (environment) errors, such as the case of encountering a slight rather than a severe sensory mismatch.

Similar to performing reaching movements, ambiguity in the cause of prediction errors

drives adaptive motor control for stabilizing motor outputs during percussion performance in which, for example, desired tones are generated by repeatedly hitting a target position on the drumhead with constant muscle strength. The structural complexity and timbral richness characteristic of many 20th- and 21st-century percussion compositions demand highly refined technical skills from performers, such as rapidly alternating between different percussion instruments varying in geometry and material properties (surface stiffness in particular), as well as switching between different hand-held mallets/sticks also varying in geometry and material properties, within the course of a piece. For example, Karlheinz Stockhausen's *Zyklus* (1959) is scored for a solo percussionist who must perform in a virtuoso manner with various mallets over an array of 21 percussion instruments ranging from marimba and suspended cymbals to high-pitched triangles. This situation poses various challenges for the nervous system to program appropriate motor parameters for different strokes: the motor system is required to demonstrate enough flexibility to adapt its motor commands to the dynamic properties of the acoustical and vibromechanical responses generated by distinct sounding objects being set into vibration.

Ambiguous prediction errors inevitably occur due to the fact that both the physical characteri stics of struck objects and the kinematic properties of striking movements can influence the perceived sensory feedback from an impact. In particular, for real impact sounds in natural scenes, both the physical stiffness coefficient of the struck object and the impact velocity of the mallet are inversely correlated with the contact duration (Chaigne & Doutaut, 1997), which consequently affects the perceived loudness and brightness of the attack portion of the impact sound (Giordano, Rocchesso & McAdams, 2010). The mechanism is clear from a physical perspective: a decrease in the contact duration between mallet and sounding object would result in a more efficient excitation of the high-frequency vibrational modes of the sounding object and, consequently, an increase in the high-frequency energy of the radiated sound. Previous psychophysical studies have shown that object properties and sound-producing actions can be

"heard" from sounds. With acoustical information alone, listeners can not only discriminate or identify the causal properties (or structural invariants) of sound-generating objects (Warren & Verbrugge, 1984), e.g., size or shape (Carello, Anderson & Kunkler-Peck, 1998; Kunkler-Peck & Turvey, 2000; Grassi, 2005; Grassi, Pastore & Lemaitre, 2013), hardness (Freed, 1990; Giordano et al., 2010) or source material (McAdams, Chaigne & Roussarie, 2004; Giordano & McAdams, 2006; McAdams, Roussarie, Chaigne & Giordano, 2010), but they can also recover veridically the causal properties of sound-generating actions (Lemaitre & Heller, 2012; Lutfi, Liu & Stoelinga, 2011) and mechanical events or scenes (Warren & Verbrugge, 1984; Cabe & Pittenger, 2000; Stoelinga, Hermes, Hirschberg & Houtsma, 2003; Houben, Kohlrausch & Hermes, 2004, 2005). Semantically reconstructing the characteristics of sound-generating actions (i.e., transformational invariants, Warren & Verbrugge, 1984) is potentially achieved via the ventral pathways wiring synaptic projections from temporal to prefrontal and premotor regions (Rauschecker & Tian, 2000; Kohler, Keysers, Umilta, Fogassi, Gallese & Rizzolatti, 2002). Because the acoustical features could be perceptually mapped onto a compound space of action- and object-related attributes, and there are substantial overlaps between the acoustical features that are determined by the action- and object-related attributes, respectively (see Appendix B for detailed acoustical analyses), ambiguity would emerge during the credit-assignment process. The working strategy that most efficiently signals a fast re-stabilization of the motor outputs for reliable sensorimotor control, is to announce a compensatory response command to the motor apparatus, so that the perturbation effects with uncertain "nature," either extrinsic or intrinsic, can be quickly counteracted (i.e., using a nonspecific adaptation strategy, Wei, Wert & Körding, 2010).

1.3 Minimum-variance optimization as a domain-general principle across multisensory processes and motor state estimation

Imagine that you are struggling to escape from a dark secret chamber. Surely your survival instincts lead you to hunt for weak spots on the walls and ground. Since the visual channel has become inaccessible to you, your urge to survive will drive the action of knocking upon the surrounding obstacles to perceptually navigate the textures of the materials through transient auditory and haptic (kinesthetic plus tactile) feedback. A question inspired from the above fictional scenario might be: Have you ever considered why your attentional weights can be assigned dynamically to the most important or "reliable" information in such a rapid fashion to make it possible to integrate optimally the available sensory information? More importantly, what is the influence of the perceptual weighting processes on our motor interaction with objects in the environment? This section will review in order the work on probabilistic modeling for perceptual estimation of object/material properties, and Kalman-filtering schemas for optimal-state estimation in motor control.

The past decade has witnessed a developing trend of incorporating probabilistic modeling into the experimental analysis of within- and cross-modal information integration. Minimum-variance optimization, of which the theoretical basis is reliability-based cue weighting, has been widely used to model the integrated perceptual processing of static features (shape, slant, texture, etc.) or spatial localization of objects in the environment. Accordingly, an optimal human observer combines task-relevant cues from multiple sources of information by taking a weighted linear average of each cue estimate (e.g., size of object) in a single sensory modality (e.g., vision or touch), in which the weight is proportional to the reliability (inverse of variance) associated with each unimodal cue. This statistically optimal prediction has found its backbone in a number of different human psychophysical paradigms, both within sensory modalities (Jacobs,

1999; Landy & Kojima, 2001; Knill & Saunders, 2003; Hillis, Watt, Landy & Banks, 2004; Toscano & McMurray, 2010) and across them (Ernst & Banks, 2002; van Beers, Wolpert & Haggard, 2002; Battaglia, Jacobs & Aslin, 2003; Gepshtein & Banks, 2003; Alais & Burr, 2004). The attractiveness of the reliability-based cue combination is that it provides a mathematically straightforward account for why precision (reliability) of a multisensory percept is maximized (Jacobs, 2002; Shams & Seitz, 2008) despite ubiquitous noise in sensory processes (Faisal, Selen & Wolpert, 2008). There are two core principles in the reliability-based cue integration model. First, the integrated (combined) estimate is a weighted average of the individual estimates with weights that sum up to unity (Cochran, 1937). Second, the combined estimate has minimum variance (i.e., maximum likelihood) precisely when the weight assigned to each cue is proportional to its reliability. In this sense, more reliable information in a given situation is given a higher weight so that it has a greater influence on the final percept. It has further been demonstrated that the optimality of reliability-based information integration also holds even when relaxing the constraint of conventionally assumed independence between different noisy sources (Landy, Maloney, Johnston & Young, 1995). A perceptual strategy that follows the weighted linear-cue combination has proven beneficial in terms of generating perceptual estimates with the lowest possible uncertainty even with non-independent sensory noise (Oruç, Maloney & Landy, 2003). For example, with binocular disparity and motion, the correlated visual cues specifying pictorial depth are based on overlapping retinal representations.

Differing from the above-mentioned computational models, the paradigms developed by Giordano et al. (2010) and McAdams et al. (2010) adopted a regression-based statistical approach to investigate unisensory (acoustical) information integration. These studies focused on the perception of sounding objects struck by hammers, with the materials of both being varied. Using a multiple rank regression framework, they quantified the perceptual weight given by each of the participants to the different acoustical features for the perceptual task. Consistent with the

optimal cue-weighting framework, a set of intriguing findings showed that the perceptual weights of the acoustic feature clusters (Principal Components, PCs) increased with the degree to which they accurately specified the sound source and with the extent to which they could be exploited, as determined by discrimination, learning and memory abilities. The construct of information accuracy proposed by Giordano et al. (2010) shares the same basic motivation with the conceptualization of cue reliability in the optimal integration framework (Knill & Saunders, 2003) and aims to measure the extent to which the acoustical information available in a learning context reliably specifies the mechanical properties of the sound source.

There has also been a great deal of work recently that applies Bayesian decision models to perceptual behaviors. Although the retinal images continuously fluctuate with saccades, an internally programmed corollary discharge counteracts the disruption of saccadic visual inputs, whereby our perception of the visual environment remains stable. "A priori preference for slower speeds" (Weiss, Simoncelli & Adelson, 2002; Stocker & Simoncelli, 2006), may be attributable to our long-term exposure to a probabilistic distribution of velocities for all objects visually encoded across a variety of general scenes in everyday experience. In perceptual tasks that require comparison of the apparent speed of two visual gratings at different contrasts (Weiss et al., 2002; Stocker & Simoncelli, 2006), lower-contrast patterns consistently appear to be slower than higher-contrast patterns (underlying drivers' tendencies to bunch up in rainy or foggy conditions with inferior visual contrast). Low-contrast patterns reduce the reliability of the received sensory cues for visual motion, so that an optimal mechanism of generating the motion estimate with the lowest possible uncertainty is to integrate the current sensory cues with a prior belief about the estimate (i.e., prior of slow-speed motion). This is exactly what has been found for the human perception of visual motion operating in a Bayesian way: noisy sensory measurements of an attribute are multiplied together with a prior probability about which world structures are most likely a priori (Stocker & Simoncelli, 2006). The prior over possible states of

the environment provides an important source of information available to maximize the reliability of a perceptual estimate. The Bayesian decision model can also explain the behavioral evidence that the perception of the forces exerted to manipulate an object relies on both sensory inputs about the actual properties of the object and probabilistic knowledge about the properties of previously manipulated objects (Körding & Wolpert, 2004a).

The Bayesian inference is an extended statistical framework of the maximum-likelihood (ML) model. The ML model assumes a flat (non-informative) prior probability density function for the statistical inference. For example, an object's size can span a wide range of values in natural contexts. These models (Bayesian or ML), as tested in previous studies, indicate that the same (variance-minimizing) schema is shared across different situations and tasks. The functioning of such a schema is to achieve perceptual optimality by integrating multiple sources of information that are associated either with different sensory cues within a single modality or across several modalities, or with prior probabilistic knowledge about a to-be-estimated parameter (a static shape/texture or dynamic motion).

The examples of optimal information integration in the previous studies referenced so far were all based on performance in perceptual tasks. In motor tasks, Bayesian inference can also be implemented by the sensorimotor control system to maximize performance by minimizing uncertainty associated with the estimates of time-varying states of the moving limbs (Körding & Wolpert, 2004a; Wei & Körding, 2010). The planning of appropriate motor commands essentially relies on the maintenance of an accurate estimate of the configuration of our body as we move in or interact with the environment. Control signals for motor output are often characterized by temporal lags. The Kalman-filtering model has been proposed for understanding how (delayed) sensory and feedforward information are averaged optimally over time to drive an efficient state estimation with minimal uncertainty (Goodwin & Sin, 1985; Wolpert et al., 1995; Deneve, Duhamel & Pouget, 2007). In essence, the Kalman-filtering model optimally estimates the

current configuration of our body (or more precisely of the action-related effector) based on its predicted state using feedforward estimates and incoming sensory feedback. This optimal state estimation can be recursively implemented to achieve reliable planning of motor commands and increase behavioral precision in the presence of motor (Harris & Wolpert, 1998) and sensory noise (Osborne, Lisberger & Bialek, 2005) in sensorimotor control problems.

The optimal strategy of information integration also appears to be implemented when people make online adjustments to movement trajectories by combining sensory feedback across multiple modalities. Humans can optimally integrate visual and proprioceptive (or vestibular) feedback from the limbs (or the body) to guide online corrections of reaching (Saunders & Knill, 2003, 2004; Sober & Sabes, 2003), spatial updating (Frissen, Campos, Souman & Ernst, 2011) or cyclical bimanual coordination (Ronsse, Miall & Swinnen, 2009). However, voluntary movements, such as walking on the floor and playing percussion instruments, usually make frequent contacts with external objects. Sensory information generated during the interactions between our moving effectors and objects can come from haptics and audition, but not necessarily vision. For example, we look straight ahead when walking, and percussionists look at the conductor or other players rather than their own hands or instruments when playing. How the processing of multimodal information generated during interaction with objects could feed back into the control of repeated actions has remained largely unexplored.

1.4 Scope of the present study

Sensory prediction errors (Miall & Wolpert, 1996; Shadmehr et al., 2010) occur when there is a discrepancy between the predicted and the observed sensory consequences of a motor command. Stabilizing dexterous strike-based interactions with sounding objects (e.g., drumming) requires efficient updating of the kinematic state (wrist movement velocity) to counteract any

sensory mismatch such as a mis-expectation of the impact properties (e.g., resistive force) based on the stiffness of the object. The prediction errors may naturally come from multiple sensory modalities, particularly audition (impact sounds) and haptics (a combination of cutaneous and kinesthetic sensations). It is unknown how audition and haptics are weighted to drive compensatory motor responses during repeated percussive events. We investigated this issue by measuring compensatory striking kinematics in a virtual-reality set-up that allowed for the alteration of both haptic and auditory feedback as sensory consequences of striking events, alone or in combination.

Previous studies have examined how sensory prediction error drives motor compensation by imposing two distinct types of disturbance effects on the sensorimotor control loop: either during the execution of movement (force-field perturbation: Shadmehr & Mussa-Ivaldi, 1994; visuomotor distortion: Tseng et al., 2007) or by altering the properties of the objects upon which repeated movements operate. The latter type in essence perturbs the touch-generated sensory feedback, so that an interactive action that was in fact accurate appears to have contained kinematic errors. Over the course of such "deceptions," compensatory motor responses are prepared for subsequent actions. Two previous studies have shown that the auditory processing of altered object-related information, i.e., material brittleness (Castiello et al., 2010) or geometry (Sedda et al., 2011) of a to-be-grasped target, which is carried by the contact sound, plays a significant role in modulating the kinematics (e.g., velocity or grip aperture) of reach-to-grasp movements. During locomotion, the unpredictable physical properties of irregular terrains have a strong modulatory effect on the compensatory gait pattern and/or lower-extremity kinematics (Ferris, Louie & Farley, 1998; Marigold & Patla, 2005; MacLellan & Patla, 2006). Consider, however, the context of another pervasive sound-generating gesture, striking objects with a tool such as a rigid probe or stylus. The mechanisms by which the integrated processing of impact-generated auditory and haptic information guides the control of striking kinematics are

not yet clear.

Psychophysical evidence exists for the acoustical influences on tactile sensations, e.g., perceived frequency content of a vibration at the skin (Gescheider, 1974; Guest, Catmur, Lloyd & Spence, 2002) or perceived quantity of tactile taps (Bresciani, Ernst, Drewing, Bouyer, Maury & Kheddar, 2005). However, haptic information appears to dominate auditory cues (echoing the postulate of "modality appropriateness", Welch & Warren, 1980, according to which sensory feedback channels are weighted based on long-term experience with respect to their efficacies of maximizing perceptual performance) across a broad range of audio-haptic processes. These include estimating the roughness of manually manipulated objects (Lederman, 1979; Lederman, Klatzky, Morgan & Hamilton, 2002; Lederman & Klatzky, 2004) and identifying walked-upon materials (Giordano, Visell, Yao, Hayward, Cooperstock & McAdams, 2012; Giordano & Avanzini, 2014). This set of evidence is in qualitative agreement with a probabilistic framework (Ernst & Banks, 2002; Gepshtein & Banks, 2003) that potentially underlies multisensory integration. In such a framework, a unified minimum-variance percept is formed by merging sensory information optimally across modalities, weighted by their modality-specific reliability (principle of maximum likelihood, or ML). The majority of past empirical evidence on the existence of such statistical decision-making strategies has been collected by focusing on purely perceptual situations in which multiple sources of information are combined to estimate external sensory attributes, neglecting the extent to which sensorimotor functioning also runs on such an optimal inference of state probabilities.

Notable exceptions to this trend are studies that have examined the effects of integrating optimally non-auditory feedback (e.g., visual, proprioceptive or vestibular) concerning the configuration of the moving limbs (or body) on early planning (Sober & Sabes, 2003) and online correction of limb movements (Saunders & Knill, 2004; Ronsse et al., 2009) or on spatial updating for self-motion (Frissen et al., 2011). However, we make frequent contacts (or

tool-based contacts) with objects when moving voluntarily. In such cases, the task-relevant sensory information that can be transformed into goal-directed motor commands (Pouget & Snyder, 2000) will include the patterns of activity in the sensory afferents innervated during the contact events (somatosensory and auditory encoding of the mechanical contact states is more direct than vision). Along these lines, one might ask whether the minimum-variance strategies also generalize to sonic-interaction tasks that involve both audition and haptics during contact with objects in the peripersonal space.

Here participants learned to maintain a target velocity of the hand-held stylus to strike repeatedly an audio-haptic virtual object synthesized when the tip of the stylus reached a given vertical position. With this set-up, impact-property feedback (acoustical and tactual) was experimentally determined not only by an intrinsic parameter—the virtual object's audio-haptic stiffness but by an extrinsic input—the striking velocity. A surreptitious perturbation to the audio-haptic stiffness of the struck object would thus confuse the estimate of the source ("limb or object?") of prediction errors in impact properties, which would drive motor corrections. In particular, the striking velocity and the struck object's stiffness would affect the acoustical properties of the synthesized impact sounds in the same direction (see section 2.3.1: Acoustical stimuli). However, they would differentially affect the two major impact-related tactual signatures encoding object hardness (i.e., surface deformation and resistive force of the struck object, LaMotte, 2000; Friedman, Hester, Green & LaMotte, 2008; see Fig. 2.2). One might thus anticipate that different participants would show opposite directions of motor compensation (by increasing vs. decreasing striking velocity) if they preferentially relied on one of the two types of altered tactual cues rather than the other, but a relatively uniform pattern (in the same direction) of compensation to the altered auditory cues (see Chapter 4: Discussion). We tested this idea by manipulating the acoustical and haptic feedback in isolation.

We also addressed the issue of modality dominance by changing the auditory and haptic

stiffnesses in combination, either congruently or incongruently (e.g., an increase in auditory hardness heard in sound and a decrease in haptic hardness felt by touch). Sensory dominance would be characterized by a potential asymmetry of crossmodal congruency effects (Driver & Spence, 2004) across audition and haptics (see section 3.1: Outline of data analyses). Furthermore, we measured modality-specific motor variability: motor variability resulting from conditions of auditory or haptic feedback alone. Given that motor variability reflects uncertainty in representation of the sensory inputs (Osborne et al., 2005; Ronsse et al., 2009) and/or central motor planning (Churchland, Afshar & Shenoy, 2006) during sensorimotor processes, we tested whether a ML account of modality-specific motor variability can predict how humans plan compensatory actions when faced with bimodal prediction errors. Expert percussionists have demonstrated superior precision in motor (Fujii, Kudo, Ohtsuki & Oda, 2009) and perceptual tasks (Lutfi et al., 2011; Cicchini, Arrighi, Cecchetti, Giusti & Burr, 2012). Musical training appears to impose plasticity effects on the neural responses to audio-tactile incongruency in a perceptual detection task (Kuchenbuch, Paraskevopoulos, Herholz & Pantev, 2014). It remains unclear, however, whether the motor-based multisensory process operates as a function of music-expertise factors, or could instead be identified as a generalized state-inference strategy that aims towards an optimality in terms of the lowest possible motor uncertainty independently of prior experience with the task (Ronsse et al., 2009). The present study also investigated this issue.

Chapter 2

Methods

2.1 Participants

Forty-two right-handed participants (26 female; mean age = 22.5 years, $SD = 4$) were included in the study. They were divided into three groups of 14: non-musicians (very limited prior musical training, $M = 0.9$ years, $SD = 1.4$), non-percussionist musicians (NP-musicians, hereafter; at least four years of musical training but not in percussion, $M = 12.6$ years, $SD = 5.1$), and percussionists (at least five years of professional training in percussion, $M = 11.6$ years, $SD = 3.7$; currently performing or practicing). All participants had normal hearing (Martin & Champlin, 2000; ISO, 2004), did not report any motor or haptic deficits, and were naïve with respect to the experimental goals. The protocol was certified by the McGill Review Ethics Board (Certificate 67–0905) and all participants gave written informed consent prior to the experiment.

2.2 Apparatus

The task was performed inside an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). While seated in a chair, each participant was instructed to hold,

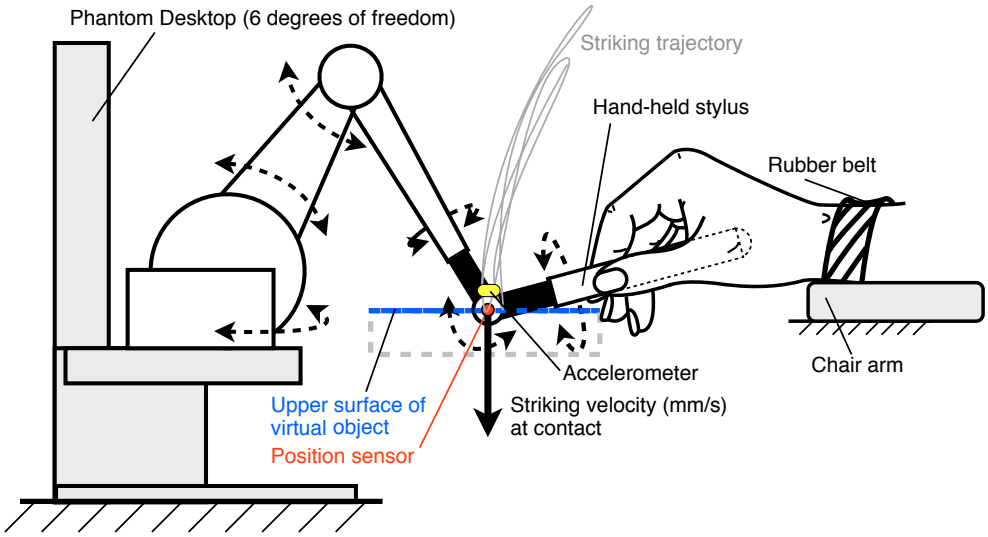

Fig. 2.1 Schematic illustration (side view) of the experimental setup: the Phantom Desktop virtual-reality haptic device. The solid blue line marks the vertical position of the upper surface of the virtual object. The solid black arrow indicates the vertical striking velocity at impact with the upper surface of the virtual object. The dashed curved arrows indicate the six degrees of freedom in which the linkage structures and the hand-held stylus can be displaced freely.

with the dominant hand (right), the stylus of the Phantom Desktop^{TM}(Geomagic Solutions, Morrisvelle, NC), a six-degrees-of-freedom robotic device used to deliver the haptic stimuli (see Fig. 2.1). The position signal of the tip of the stylus was measured at 1kHz by optical encoders at the joints of the linkage structure connected to the stylus. Velocity was computed from position data using numerical differentiation along the vertical axis. This haptic device was interfaced with a real-time Pd (version 0.42.5)[1] program for the synthesis of impact sounds, and with a custom C++ program for the control of haptic feedback (see section 2.3: Stimuli, below). Sound stimuli were amplified with a Grace Design m904 stereo monitor controller (Grace Digital Audio, San Diego, CA) connected to the optical port of a Windows workstation used to control the experimental variables, and were presented binaurally through Sennheiser HD280

[1]M. Puckette, PureData (Pd), http://www.puredata.org, accessed 6 April 2014.

headphones (Sennheiser Electronics GmbH, Wedemark, Germany). Sound pressure level (SPL) of acoustical stimuli was measured offline prior to the experiment using a Brüel & Kjær Type 2205 sound-level meter (A-weighting) coupled with a Brüel & Kjær Type 4153 artificial ear (Brüel & Kjær, Nærum, Denmark). Verbal instructions about the task and visual feedback on the appropriateness of the participants' performance were presented via the computer monitor (for details see section 2.4: Procedure, below).

The haptic device was stabilized on a table. The participant's right forearm (except the hand and wrist) rested on the chair's right arm, which was of the same height ($\sim$ 15 cm above the table) as the upper surface of the virtual object (see section 2.3: Stimuli). The right wrist was prone and lightly restrained with a rubber belt to the chair's right arm so that only wrist flexion and extension could be easily used to displace the stylus. The joint angle between the upper and lower arm was adjusted to be approximately 90°. Participants were denied vision of the hand, wrist, forearm, and stylus.

The equipment for measuring impact-generated acceleration (indexing resistive force; Fig. 2.2c, d) of the stylus was a PCB Piezotronics (Depew, NY) 352C42 miniature accelerometer (sampling frequency = 48 kHz) connected to a Model 480E09 PCB signal conditioner, itself connected to a USB-6218 acquisition card (National Instruments, Austin, TX) for analog-to-digital conversion. The accelerometer was adhesively mounted with wax on the top surface of the stylus gimbal.

2.3 Stimuli

Audio-haptic stimuli were synthesized impact sounds and haptic feedback simulating the sensory outcome of stylus-based strikes upon a static sounding object (Avanzini & Crosato, 2006; Itkowitz, Handley & Zhu, 2005). To impose perturbing effects on the struck object, we

manipulated its mechanical parameter—stiffness coefficient K. From the haptic standpoint, K establishes the relation between the object's degree of deformation (surface compression) and the perceived resistive force, and its value is determined by the Young's moduli and Poisson ratios of both solids in contact (Chaigne & Doutaut, 1997). From the auditory standpoint, higher values of K generate an increase in the subjective hardness of the struck object(s) perceived from the synthesized impact sounds (Giordano et al., 2010).

2.3.1 Acoustical stimuli

Impact sounds (44.1 kHz, 16-bit resolution; peak intensity = 75 dB SPL) were synthesized using a real-time physically inspired model of an ideal struck bar with five independent vibrational modes and an internal dissipation (Avanzini, Rath, Rocchesso & Ottaviani, 2003; Avanzini & Crosato, 2006). Communication latency resulting from the sound synthesis was on the order of a few milliseconds, which is well below the typical experimental estimates for the temporal window of auditory-tactile integration (around 200 ms; Bresciani et al., 2005). A previous study by Avanzini & Crosato (2006) using the same experimental setup reported that no participants perceived any kind of noticeable intermodal latency. See Appendix A for detailed algorithms of the synthesis model. The resonant frequency of the lowest vibrational mode, F, was set to 100 Hz. Higher-order modal frequencies were tuned according to the most prominent resonances of a bar clamped at one end and free at the other (Fletcher & Rossing, 1991), i.e., they were multiples of F by $\{6.26, 17.54, 34.37, 56.82\}$. The properties of the synthesized acoustical signal were determined in real time by two input parameters—acoustical stiffness coefficient (acoustical K; see Fig. 2.2a) and striking velocity at impact (Fig. 2.2b). Both the acoustical K and the impact velocity are inversely correlated with the contact duration (Chaigne & Doutaut, 1997) in the model, which consequently affects the acoustical features, particularly the loudness and brightness of the attack portion of the impact

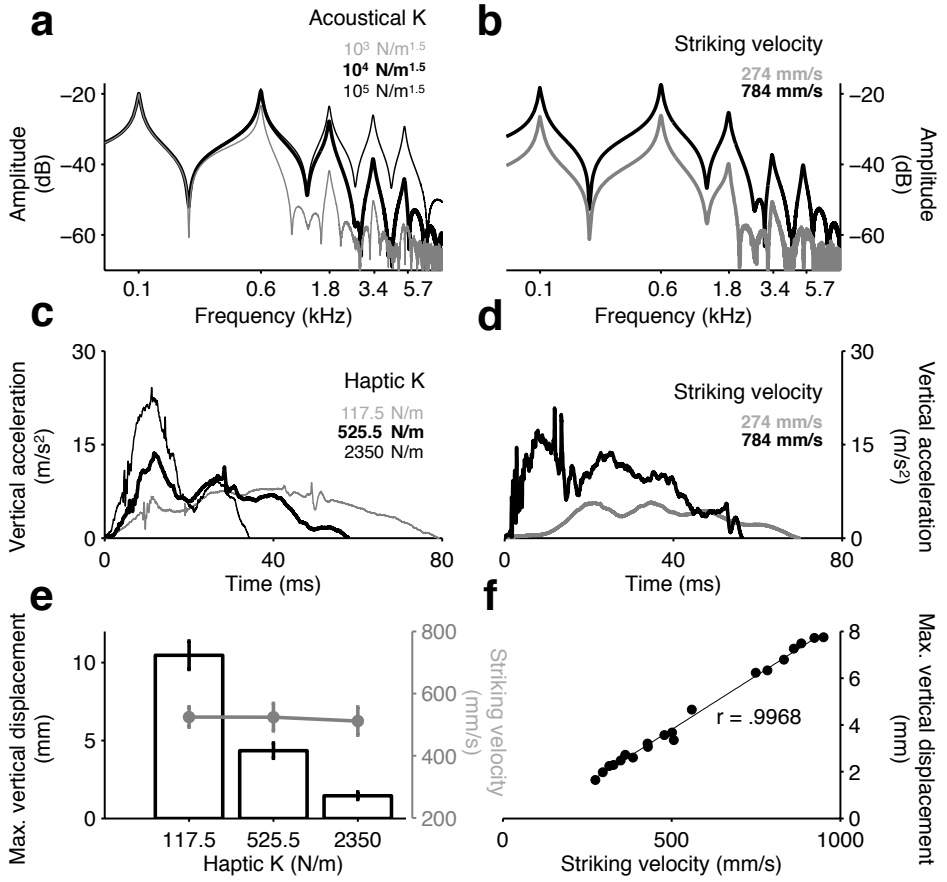

Fig. 2.2 Properties of audio-haptic stimuli as a function of stiffness coefficient K and striking velocity at impact. **a-b** Spectra of impact sound as a function of either acoustical K (**a** synthesized with constant impact velocity of 500 mm/s) or striking velocity (**b** acoustical $K = 10^4$ N/m$^{1.5}$). Vibrational modes with periods longer than contact duration fail to be excited efficiently; either larger acoustical K or higher striking velocity leads to shorter contact duration. **c-d** Force-related cue, vertical acceleration (stylus-object contact starts at $t = 0$) of the stylus as a function of either haptic K (**c** equivalent striking velocities, $M = 491$, $SD = 16$ mm/s) or striking velocity (**d** haptic $K = 525.5$ N/m). **e-f** Deformation cue, maximal vertical displacement (black bars) of the stylus relative to the virtual object surface decreases with increasing haptic K (**e** equivalent striking velocities across haptic Ks, one-way ANOVA $p > .96$), but increases with increasing striking velocity (**f** haptic $K = 525.5$ N/m). In **e**, grey points represent the mean stylus velocities used to strike the three different haptic surfaces. Error bars indicate $\pm$ 1 SEM ($N = 20$ strikes).

sound (Giordano et al., 2010). A decrease in the contact duration would result in a more efficient excitation of the high-frequency vibrational modes of the sounding object and, consequently, an increase in the high-frequency energy of the radiated sound. We quantified the extent to which there are overlaps between the acoustical features that are influenced by a user's striking velocity (action related) and the model's acoustical K (object related) (Appendix B). The overlaps proved to be substantial for the spectrotemporal features of the impact sound that perceptually specify a struck object's stiffness and the properties of impactor/object interaction (Giordano et al., 2010; McAdams et al., 2010), which provides a testing ground for the participants' compensation of their motor output in response to the perceived changes in the sounding object's properties.

In this study, acoustical K could vary across five log-spaced levels $\{10^3, 10^{3.5}, 10^4, 10^{4.5}, 10^5\}$ N/m$^{1.5}$ centered around a baseline level of 10^4 N/m$^{1.5}$.

2.3.2 Haptic stimuli

Impact-related touch feedback was simulated using a dissipative contact model when the head (gimbal) of the hand-held stylus reached the surface of the virtual object (15×15 cm surface area). A reaction torque generated by the motors inside the device was delivered onto the linkage structures connected to the stylus, simulating a vertical resistive force proportional to the instantaneous normal displacement of the stylus gimbal relative to the virtual object's upper surface. The force was determined by a linear combination of the gimbal's displacement and velocity, weighted by a haptic stiffness coefficient (haptic K) and a dissipative (damping) factor, respectively (Appendix A). The inclusion of the dissipative component could alleviate unstable (oscillatory) feedback (Colgate & Brown, 1994). Both the haptic K and the striking velocity at impact could modify the haptic feedback, but differentially for the force-related and displacement-related features (see Fig. 2.2c–f). Haptic feedback was programmed with the OpenHaptics^{TM} Toolkit (Itkowitz et al., 2005). In this study, haptic K could assume one of five

log-spaced values {117.5, 248.5, 525.5, 1111.2, 2350.0} N/m (baseline = 525.5 N/m). The largest stiffness coefficient here corresponds to the maximum stiffness (along the vertical direction) that can be achieved by the Phantom Desktop.

2.4 Procedure

A schematic that depicts the experimental design and sequence is presented in Fig. 2.3. Each trial was divided into three consecutive phases. During an initial training phase, participants learned to constrain their striking velocities of the downswings within a target range (430–570 mm/s). They received on-screen feedback on a strike-by-strike basis: “too slow”, “correct” or “too fast” was displayed immediately after a stroke if the velocity at impact was below, within or above that range, respectively. This phase ended when a participant had produced five consecutive strikes with the “correct” velocities. During a subsequent maintenance phase, participants continued striking with the same trained target velocity in the absence of feedback on correctness. This phase ended after five strikes independently of whether they were within or outside the target velocity range. The stiffness coefficient K of the simulated object was fixed to the baseline value across these two phases. At the beginning of a final change phase (without feedback on correctness), the value of the acoustical and/or haptic K could be suddenly altered, alone or in combination, depending on the experimental condition. Participants were told to continue striking with the same trained velocity and to ignore any changes in the properties of the object. This last phase ended after 20 strikes. Participants were required to strike with a tempo (no faster than three strikes/second) of their choice, provided that they kept it unchanged throughout an entire trial.

We investigated three conditions. In the haptic-only condition (100 trials), participants were presented with the simulated haptic object and with a continuous white-noise auditory

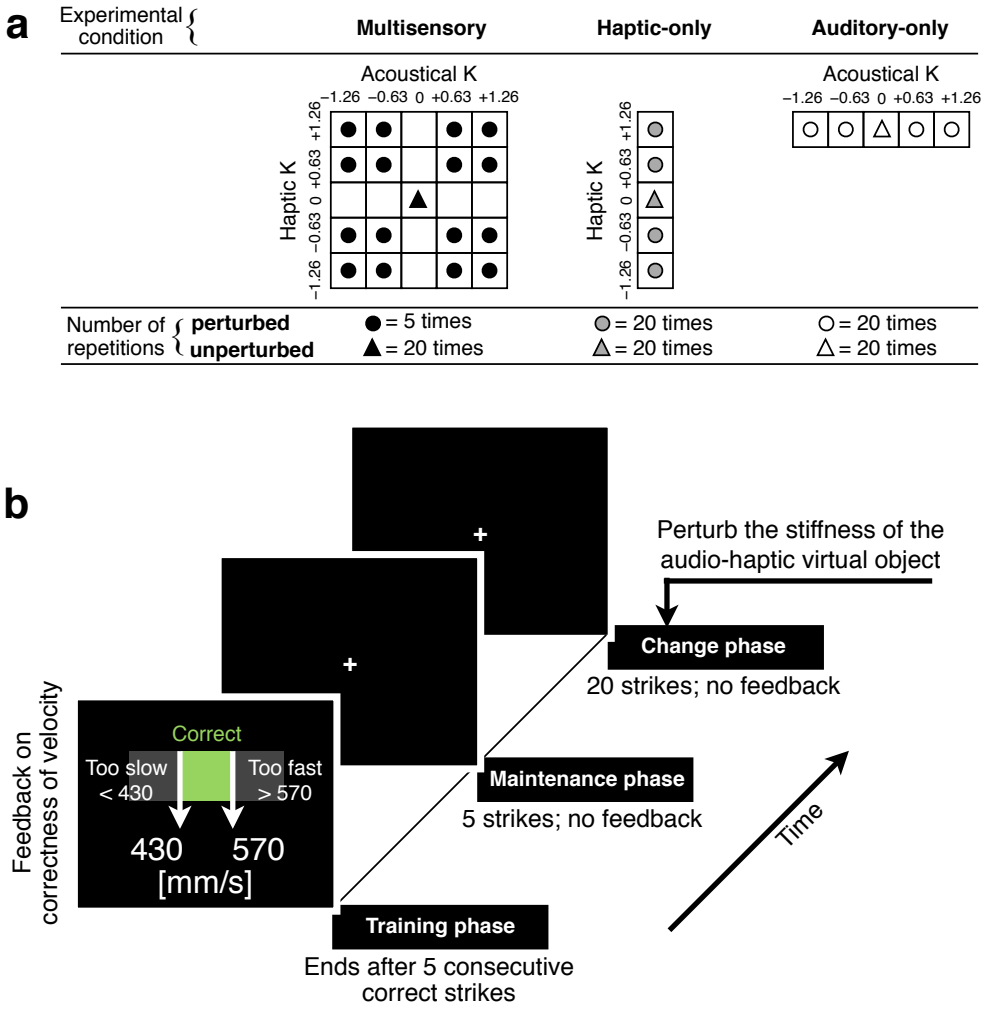

Fig. 2.3 **a** Experimental conditions. Circles and triangles indicate the change-phase stimulus levels (K values are log transformed and standardized) of the perturbed and unperturbed trials, respectively. **b** The time course of one trial, showing the 'training,' the 'maintenance' and the 'change' phases.

masker presented at 75 dB SPL. In the auditory-only condition (100 trials), when the stylus reached the contact position, participants were presented only with the synthesized impact sound but no impact-related haptic feedback. During the audio-haptic (bimodal) condition (100 trials), both the impact-related haptic and acoustical feedback were available. During the change phase of a bimodal trial, the acoustical and haptic K of the simulated object could assume either the

baseline value (unperturbed trial: triangles in Fig. 2.3a) or any of the 16 factorial combinations of the non-baseline values (perturbed trial: circles in Fig. 2.3a). Bimodal perturbations also depended on crossmodal congruency. In the congruent condition, both the acoustical and haptic K either increased or decreased relative to the baseline level. In the incongruent condition, the acoustical and haptic K changed in opposite directions (acoustical K increased whereas haptic K decreased, or vice versa).

Each participant completed 20 blocks of 15 trials each for a total of 300 trials in three sessions of approximately 1.5 hours each on different days. Each block was subdivided into five subblocks during which participants were presented with three trials of different conditions (random order of conditions within each subblock). The 15 trials per block comprised five auditory-only trials (different change-phase acoustical K values), five haptic-only trials (different change-phase haptic K values), and five audio-haptic trials (one unperturbed trial and equal number of congruent and incongruent trials). Throughout all the audio-haptic trials, each of the possible combinations of the change-phase non-baseline haptic and acoustical K values was repeated five times, whereas the baseline value of the haptic/acoustical K was repeated 20 times. This design thus balanced each of the five K levels of the two dimensions (acoustical vs. haptic) with an identical number of trials for the unimodal and bimodal conditions, e.g., the haptic K of 117.5 N/m occurred equivalently 20 times for the haptic-only condition and the audio-haptic condition.

Chapter 3

Results

Data were analyzed with three linear mixed-effects models (LMM, West, Welch & Galecki, 2006) fitted using the SAS® PROC MIXED routine (Littell, Milliken, Stroup & Wolfinger, 2006). The LMMs included both fixed effects (constant parameters) and random effects (assumed to follow a normal distribution) measuring the average behavioral trend (mean linear motor effect of different levels of stiffness perturbation) and its inter-individual variability within the population of interest, respectively. The results of the LMMs were illustrated by the normal cumulative distribution functions (CDFs, Fig. 3.1b), with the fixed-effect estimates corresponding to the horizontal-axis values for which the CDFs reach the probability level of 0.5 and the random-effect estimates corresponding to the variances of the CDFs. Model-selection procedures in terms of the specification of the variance-covariance matrix of residuals and the significance tests for random effects followed a top-down strategy (West et al., 2006, pp. 39–41) based on likelihood-ratio statistics. In situations in which two random effects were considered in a model and the significance of each of them needed to be tested, $\chi^2(1,2)$ was used to denote the test statistic of which the asymptotic null distribution was a mixture of $\chi^2(1)$ and $\chi^2(2)$ distributions, with each having an equal weight of 0.5 (Verbeke & Molenberghs, 2000). We

adjusted p values with the Bonferroni correction for multiple post hoc comparisons.

3.1 Outline of data analyses

The first LMM analyzed the effects of perturbing the stiffness coefficient K on the change-phase striking velocities in the two unimodal conditions—auditory-only (no impact-related haptic feedback) vs. haptic-only (with auditory masking noise). The sudden perturbation of the struck virtual object's stiffness would change simultaneously but oppositely the magnitude of the object's surface deformation and the magnitude of compressional force produced by the stylus against the object. The motor compensation for the altered deformation vs. force cues being preferentially attended to by different participants was expected to be in opposite directions in that increasing (decreasing) striking velocity would increase (decrease) the magnitude of both cues. Taking advantage of its merit in detecting and estimating variability in the motor effects of stiffness perturbation, the LMM was used to classify the 42 participants based on the direction of their motor compensation for the haptic perturbation.[1]

With the second LMM, we aimed to investigate the existence of sensory dominance by examining potential asymmetries of crossmodal congruency effects across audition and haptics. Here, the crossmodal congruency effects (the consequence of a contrast between incongruent and congruent conditions, see Driver & Spence, 2004) can index the extent to which a motor effect of

[1]Note that compensation doesn't necessarily imply adaptation in the present study. In classic perturbation paradigms, sensory feedback is perturbed after a baseline phase. The magnitude of compensation is calculated by the difference between the average motor variable during the last few movements from the perturbation phase and the average motor variable of the baseline phase. In studies of speech motor control (Mitsuya, MacDonald & Munhall, 2014; Villacorta et al., 2007; Perkell, 2012), the measured variables are acoustic parameters such as formant frequencies, voice onset time, etc. Adaptation usually refers to the temporary persistence of the compensatory behavior after the removal of the sensory perturbation, indicating a temporary modification of the feedforward commands. Therefore, adaptation is often quantified by the 'aftereffect' during a post-test phase. In our recent follow-up study involving a post-test phase (10 strikes) in which the stiffness of the audio-haptic virtual object returned to the baseline value, a transient persistence of velocity compensation was clearly pronounced, indicating the evidence of adaptation (unpublished observations).

one sensory perturbation is affected by a simultaneously presented "distractor" (i.e., incongruent) perturbation in another sensory modality. If an evident intersensory asymmetry occurs while participants integrate the stiffness-related auditory and haptic prediction errors, then the motor compensation for the prediction errors in the dominant sensory modality should show less crossmodal congruency effects than the motor compensation for the prediction errors in the other modality.

Within the final LMM, we examined whether the results of potential sensory dominance revealed by the second LMM would be predictable from a ML account of sensory weighting based on modality-specific motor variabilities (noises). Through "sensorimotor transformation" (Pouget & Snyder, 2000), feedback-driven motor commands for controlling the striking velocity could be converted from both proprioceptive signals and action-related spectrotemporal features carried by the impact sounds/vibrations. Given the task of maintaining striking stability, the reliability of a source of action-related sensory feedback could be reflected by the extent to which motor noise (σ) is minimized in the presence of that sensory feedback. Consider the time period of a stylus-object contact, an estimate of the current motor signal for controlling striking velocity could originate from an internal estimator (Wolpert et al., 1995; Kawato, 1999; Wolpert et al., 1998), which integrates an efference copy of prior control signals with the feedback-driven motor commands. The signal components are approximated by independent Gaussian probability density functions with different variances (σ^2, inverse of reliability; Landy, Banks & Knill, 2011). An optimal motor-planning strategy would potentially combine the redundant control signals weighted by their relative reliabilities. If so, the reliability of a crossmodally combined estimate is maximized by summing up the reliabilities associated with the individual signals. The auditory- and haptic-only condition omitted one sensory source for the motor estimate, associated with the impact-generated haptic and acoustical feedback, respectively, compared with the bimodal contexts where a full combination of control signals was available. According to a ML

schema, the reliability ($r = \sigma^{-2}$) of a control signal that is exclusively based on the impact-related acoustical (i.e., r_{Aud}) and haptic (r_{Hap}) feedback can be estimated by subtracting the inverse of the haptic-only motor variance and the inverse of the auditory-only motor variance from the inverse of the bimodal motor variance, respectively (other signals are computationally cancelled out by the subtraction). Thus the ML-predicted audio-haptic cue weighting, W_{Aud}/W_{Hap}, can be computed using r_{Aud}/r_{Hap}.

3.2 Stiffness perturbations prompted motor compensations in auditory- and haptic-only conditions

For each trial, we considered the average change-phase velocity across the last 19 strikes because the first could not reflect the motor effect of the perturbed sensory feedback. These trial-specific motor measures were then collapsed into five means corresponding to the five different levels of the stiffness coefficient K per unimodal condition (see Chapter 2: Methods), yielding ten data points (five for auditory-only; five for haptic-only) for each of the 42 participants.

We examined the fixed effects of K (modeled as a continuous variable), condition (auditory- vs. haptic-only), and musical expertise (non-musician, NP-musician, percussionist), and of all the possible interactions between these factors. Log-transformed values of haptic and acoustical K were standardized (z-scored), resulting in five linear-spaced levels from -1.26 to 1.26 (zero-centered). Meanwhile, we included in this LMM significant random effects of both the participant-specific slope coefficients associated with K and the participant-specific intercepts [$\chi^2(1,2) \geq 15.1,\ ps < .001$]. We adopted a participant-specific compound-symmetry structure for the variance-covariance matrix of residuals specified for auditory- and haptic-only conditions heterogeneously [better fit than assuming constant residual variance, $\chi^2(1) = 229.5, p < .0001$].

This LMM explained 73.2% of the total variance (unadjusted R^2) in the striking velocity.

Overall, the participant-specific slope coefficients of K, which represent the motor effects of stiffness perturbation, were significantly different from zero [$F(1,78) = 51.1,\ p < .0001$]. A three-way $K \times$ expertise $\times$ condition interaction was marginally significant when all expertise groups were considered [$F(2,291) = 2.84,\ p = .057$], but was significant if the factor of expertise was recoded a posteriori in order to contrast percussionists with non-percussionists [$F(1,291) = 5.49,\ p = .019$]. In particular, the effect of haptic K was modulated by the percussion expertise [$F(1,291) = 4.09,\ p = .044$ for the percussion-expertise-related contrast within the haptic-only condition]: ten percussionists (71%; eight of whom were mallet percussionists, see Chapter 4: Discussion) increased their striking velocity for higher values of haptic K (stiffer haptic surface), whereas both NP-musicians and non-musicians tended to strike with lower velocity to compensate for higher values of the same variable. Only three NP-musicians (21%) and six (43%) non-musicians were estimated to have positive linear coefficients for the haptic K. The mean fixed effect of haptic K was estimated to be -17.2 for non-musicians and -18.5 for NP-musicians, whereas percussionists had a positive (7.87) average estimate for the same fixed effect (Fig. 3.1a, b). The perturbing effect of acoustical K was instead consistent across the three groups [$F(2,291) < 1$ for the post hoc contrast within the auditory-only condition]. An increase in acoustical K led to a decrease in striking velocity in all participants (average fixed-effect estimate of acoustical K was -15.24, $SD = .58$). A post-experiment inquiry revealed that participants did not feel they deviated from the target striking velocity, although most of them did perceive the changes in object properties.

In addition, a larger degree of inter-individual variability emerged for the motor effects of haptic K than for those of acoustical K, based on a between-modality comparison for the random-effect variance (haptic: 1074.4 vs. acoustical: 40.6, corresponding to the different spreads of the K-slope distributions in Fig. 3.1b).

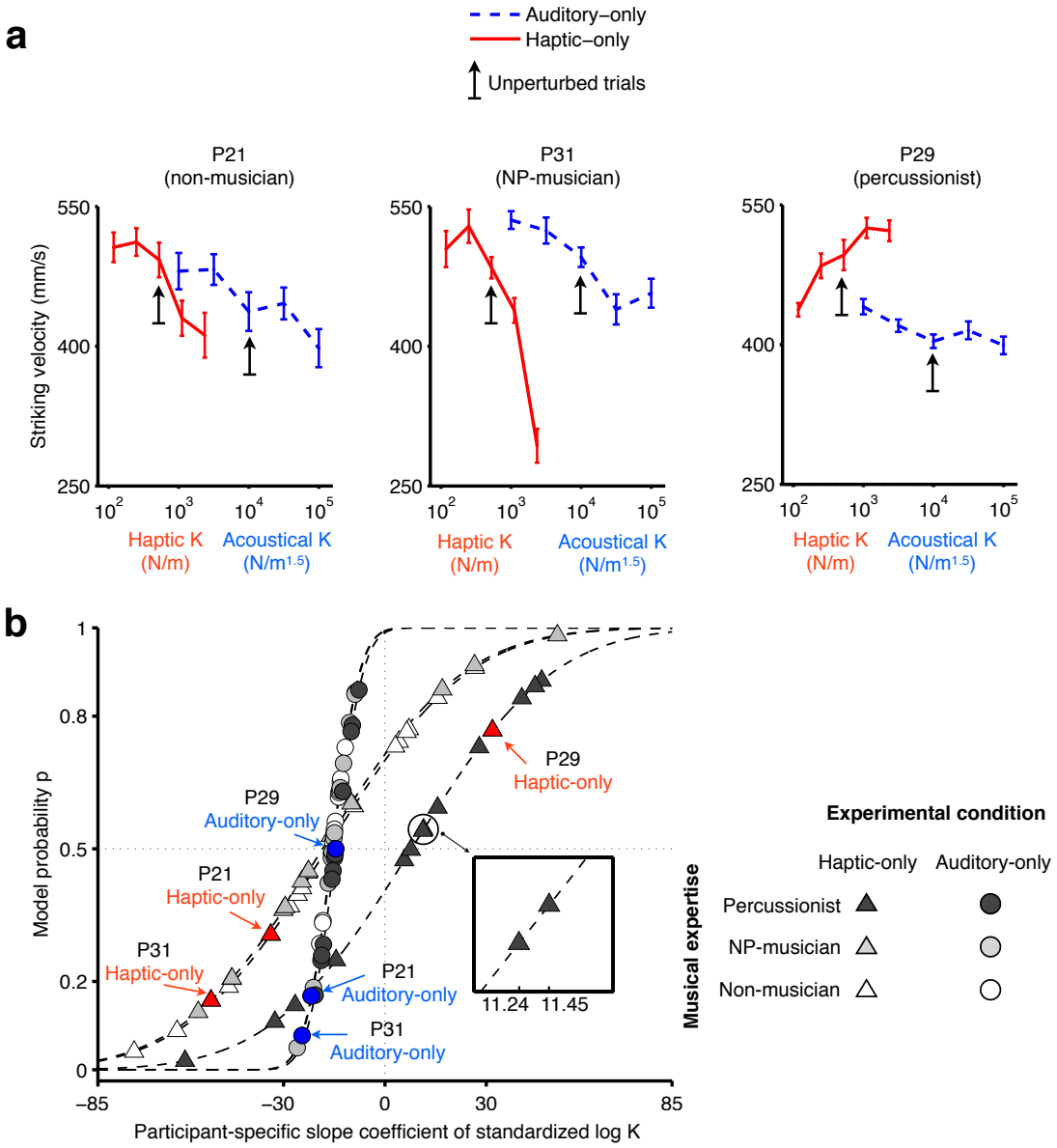

Fig. 3.1 Effects of perturbed stiffness coefficient K on striking velocity in unimodal conditions. **a** Three typical participants with different musical experience (non-musician P21, NP-musician P31, and percussionist P29). Mean change-phase striking velocity (mm/s) is plotted as a function of haptic (N/m) and acoustical (N/m$^{1.5}$) K in two different unimodal conditions (red: haptic-only; blue: auditory-only). Upward pointing arrows indicate the data of unperturbed trials (constant K). Error bars represent $\pm$ 1 SEM ($N = 20$ trials). **b** LMM estimates of the participant-specific linear effects (fixed plus random) of the stiffness coefficient K on striking velocity. Fixed-effect estimates correspond to the x-axis values for which the Gaussian cumulative distribution functions (CDFs, dashed curves) reach the probability level of 0.5; random-effect estimates correspond to the variances of the CDFs. Symbol color and shape corresponds to expertise and condition, respectively. Specific K-effect estimates for the typical participants (see **a**) are identified for illustrative purposes.

Participants produced, on average, higher striking velocities in the haptic-only condition than in the auditory-only condition [$F(1,291) = 23.8,\ p < .0001$]. This result was due to the fact that the average change-phase striking velocity of the unperturbed haptic-only trials (518 mm/s) was significantly higher than that of the unperturbed auditory-only trials (485 mm/s), [$t(41) = 4.51,\ p < .0001$]. The average change-phase striking velocity of the unperturbed trials was not of primary interest in this study. A correlation analysis revealed that the absolute value of the participant-specific linear motor effect of different levels of stiffness perturbation was independent of the average striking velocity of the unperturbed trials for both unisensory conditions, [$\rho \leq .13, p \geq .12$]. Other fixed effects of condition $\times$ expertise, $K \times$ condition, $K \times$ expertise, and expertise per se, were not significant [$F \leq 1.73,\ p \geq .18$].

In summary, this LMM revealed negative participant-specific estimates for the motor effect of acoustical perturbation. However, it revealed 19 positive and 23 negative participant-specific estimates for the motor effect of haptic perturbation, for two *classes* of individuals who were observed to compensate for an abruptly stiffer haptic object by increasing and decreasing the striking velocity, respectively. From now on, we refer to these groups as *haptic positive* and *haptic negative*, respectively.

3.3 Crossmodal congruency effects were asymmetrical between audition and haptics

We adopted the idea of contrasting the behavioral effects of crossmodally congruent with incongruent conditions to investigate sensory dominance. This idea has been used to reveal an asymmetry regarding the significant modulation of the perception of the direction of auditory motion by visual motion, but the nearly null effects of auditory motion direction on the perception of the direction of visual motion (Soto-Faraco, Spence & Kingstone, 2004b). The

participants' task in their study was to judge the direction of motion in one modality while attempting to ignore the stimuli presented in the other modality. In our present analyses of the data in bimodal conditions, we examined the extent to which the motor effects of the perturbation in a more dominant modality could prevail in cases of bimodal perturbation so that only the behavioral effects of the stiffness perturbation in the other modality were modulated by congruency. We determined the motor effect of perturbation in one modality while averaging the change-phase striking velocities at different levels of perturbation in the other modality. For example, the motor effect of the four different levels of acoustical perturbation in the congruent condition (see "Acoustical K, Congruent" in Fig. 3.2) was determined by averaging the change-phase striking velocities at the two levels of congruent haptic K associated with each of the four acoustical K levels. We aimed to examine the extent to which the sensory feedback from one modality was processed differently depending on whether it was congruent with the information from the other modality or not. The compensatory motor responses to the perturbations in the dominant sensory modality should show less crossmodal congruency effects than the reverse case.

We considered the average change-phase striking velocities (last 19 strikes) from the bimodal perturbed trials (i.e., an increase or decrease in K from its baseline level). For each participant, the mean striking velocity was taken for each of the 16 possible combinations of the non-baseline haptic and acoustical K levels. Critical to this LMM, these 16 measures were collapsed according to a 2×2 factorial contrast between modality and congruency, yielding four independent data vectors (congruent vs. incongruent, per modality) of four data points each (see Fig. 3.2). We examined in this model not only the fixed factors of standardized log K, sensory modality (auditory vs. haptic), and crossmodal congruency, but also the participant-level covariates of musical expertise and class (haptic positive vs. haptic negative), as well as the possible interactions among these factors. Meanwhile, we kept in the LMM significant random

effects of both the participant-specific slope coefficients associated with the factor K and the participant-specific intercepts [$\chi^2(1,2) \geq 61.7,\ ps < .0001$].

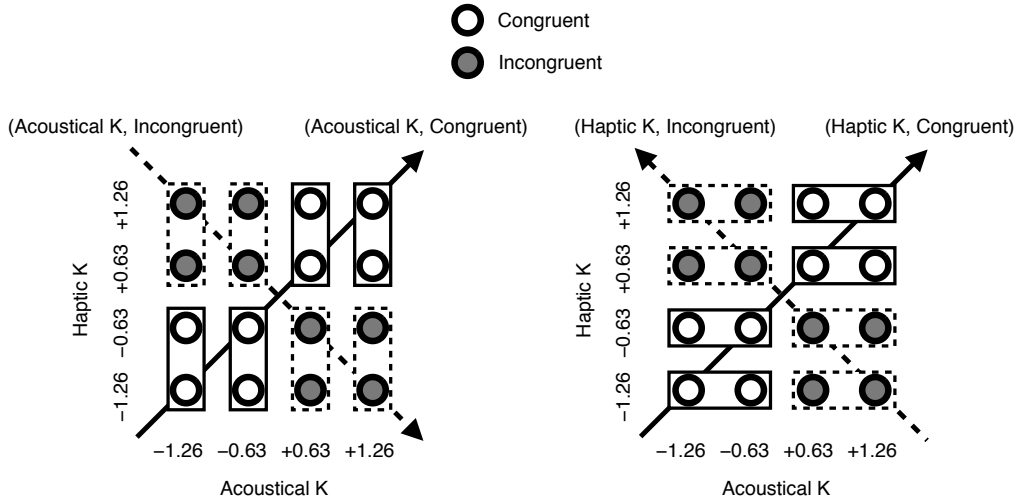

Fig. 3.2 Collapsing data into four vectors for the second LMM according to modality × congruency. K values are log transformed and standardized.

This model accounted for 87.6% of the total variance in the input data. It revealed significant fixed effects of K [$F(1,80) = 5.38,\ p = .023$], congruency [$F(1,496) = 4.02,\ p = .046$], K × congruency [$F(1,496) = 20.1,\ p < .0001$], and K × modality [$F(1,80) = 18.3,\ p < .0001$], but not of modality itself [$F(1,80) < 1$]. Critical to this analysis, there was a significant three-way K × congruency × modality interaction [$F(1,496) = 29.7,\ p < .0001$]. A pairwise between-modality contrast (haptic minus auditory) for the congruency-related absolute changes in the estimated linear coefficients of K revealed that the altered haptic feedback dominated the altered acoustical feedback in determining the motor compensation [$t(496) = -4.90,\ p < .0001$]. In particular, the population-average estimate for the motor effect of haptic K was shifted to a lesser degree by the crossmodal congruency (an asymmetry of the congruency effects in favor of haptics), compared with that of acoustical K. The mean congruency-related absolute change in the estimated linear coefficients of haptic and

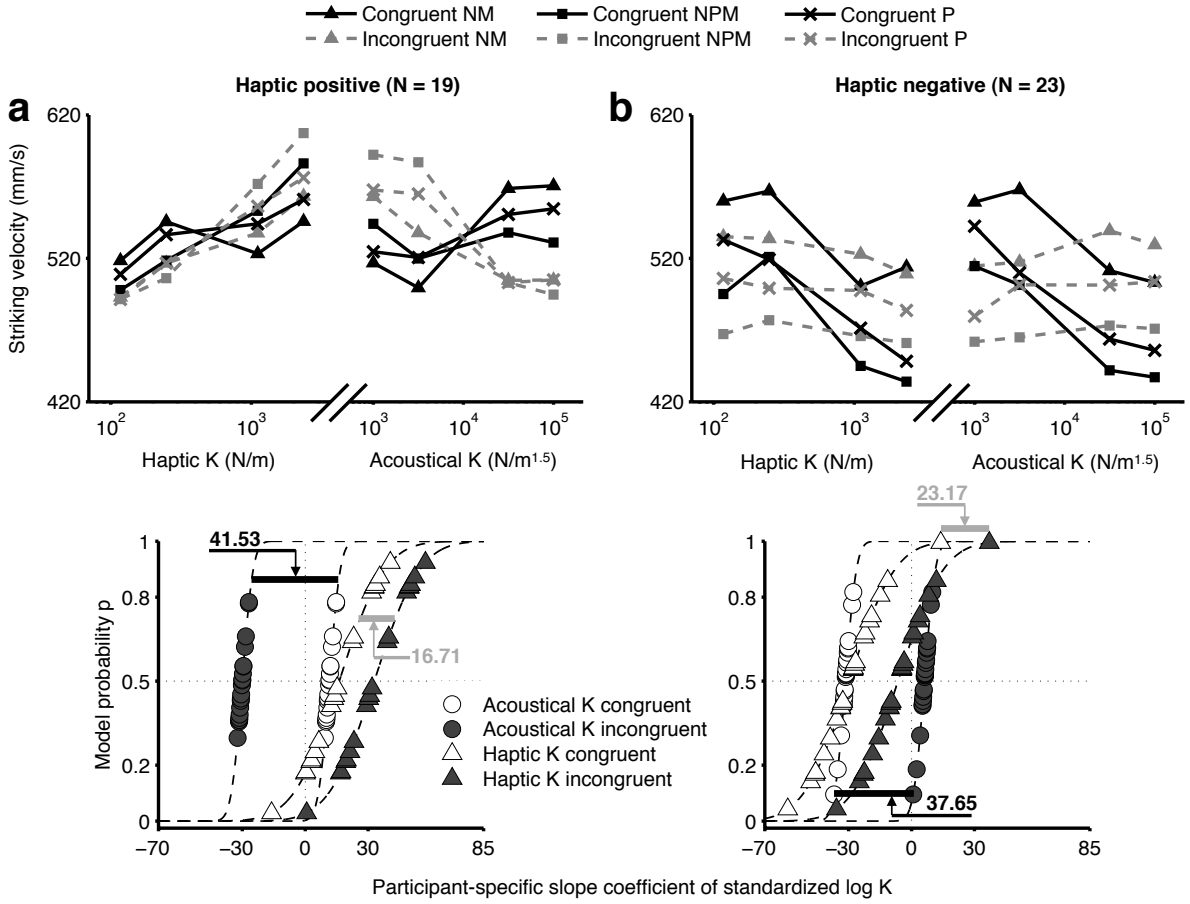

Fig. 3.3 Upper panels: across-participant mean of change-phase striking velocity in perturbed audio-haptic trials. *Note*: NM = non-musician, NPM = NP-musician, P = percussionist. Lower panels: participant-specific LMM estimates of the effect of K as a function of modality and crossmodal congruency. Horizontal black and grey bars (and numbers associated with them) indicate the absolute shifts in model estimates of the fixed effects of acoustical K and haptic K, respectively, by crossmodal congruency. Participants belonging to different classes (see texts) are distinguished by **a** and **b**.

acoustical K was 20.2 and 39.6, respectively.

Furthermore, the asymmetry of crossmodal congruency effects between audition and haptics was more pronounced for the haptic-positive individuals than for the haptic-negative individuals. We used the population-average estimate for the congruency-related absolute change in the effects of K as a measure for deriving the relative weights of the two sensory modalities.

For example, the relative weight of haptic feedback for the haptic-positive participants was $41.53/(41.53+16.71) = 0.71$ (Fig. 3.3a, lower panel). Overall, the haptic-positive individuals gave greater weight to the haptic inputs (71%) than to the auditory inputs (29%), which underpinned their stronger response bias towards haptics when compared with the relative weighting of haptics and audition (62% vs. 38%) in the haptic-negative individuals.

In addition, the $K \times$ class effect was significant [$F(1,496) = 47.9,\ p < .0001$], with the average linear coefficient of K for the haptic-negative participants being estimated to be 39.5 ($SE = 6.58$) units less than that of the haptic-positive participants. The class effect per se was significant [$F(1,496) = 10.3,\ p = .0014$], with the mean change-phase striking velocity of the haptic-negative participants being 43.4 mm/s ($SE = 22.8$) lower than that of the haptic-positive participants, which was likely due to a difference in their baseline striking velocity prior to the change phase. The test also revealed significant fixed effects of $K \times$ modality $\times$ class and $K \times$ congruency $\times$ class interactions [$F(1,496) \geq 34.8,\ ps < .0001$]. To more closely inspect these, we compared the model estimate for the motor effect of K that was associated with each modality and with the congruent and incongruent conditions separately for each class.

For the haptic-positive individuals who tended to increase their striking velocity for an unexpectedly stiffer haptic surface that was presented alone (haptic-only), they were characterized by the same motor-compensation strategy, independently of the crossmodal congruency, when receiving simultaneously the acoustical and haptic perturbations (Fig. 3.3a). In particular, compared to the unimodal haptic perturbation, the estimated linear coefficient of K was also positive (regardless of congruency). This finding provides additional evidence that reinforces the conclusion that haptic dominance is inferred from the congruency-effects asymmetry. The positive motor effect of haptic K was larger for the incongruent (32.9) than for the congruent (16.2) trials, indicating that a stiffer haptic object combined with a softer impact sound (or vice versa) could induce an enhancement in the extent of motor adaptation. The motor

effect of acoustical K, which was originally negative (e.g., striking velocity was decreased for 'harder' impact sounds) when the acoustical feedback was presented alone (auditory-only), turned out to be reversed by the positive motor effect of haptic K in the congruent audio-haptic condition.

For those haptic-negative participants who adapted to a stiffer haptic surface by decreasing the striking velocity in auditory masking noise (haptic-only; see Fig. 3.3b), adding an incongruent task-relevant acoustical perturbation appeared to affect more strongly their compensation for the altered haptic feedback in the bimodal context, compared with the haptic-positive participants mentioned above. The motor effect of haptic K estimated in our LMM for the incongruent audio-haptic trials, albeit still negative, was quite small (-6.52) on average if compared with that for the congruent trials (-29.7). The haptic dominance was still pronounced here, as revealed by a somewhat positive estimate (6.14) for the motor effect of acoustical K in the incongruent bimodal condition (the preceding LMM has revealed negative motor effects of perturbing the same variable during the auditory-only trials).

Finally, the effect of musical expertise was not significant [$F(2,496) = 1.50,\ p = .22$], but it was modulated by the effect of class [$F(2,496) = 3.16,\ p = .043$], with significantly lower striking velocity on average for the NP-musicians compared to non-musicians, but only within the haptic-negative class of participants [$t(496) = -3.78$, $p = .0012$; $|t(496)| \leq 1.71$, $p \geq .53$ for the other contrasts between expertise-related groups]. Notably, any other expertise-relevant effects failed to reach statistical significance [$F(2,496) < 1$], indicating that the above-mentioned motor effect of K and its interactions with other factors (i.e., modality, congruency, class, or interactions between these) were generalizable across the three expertise-related groups.

3.4 Sensory dominance in compensatory motor behavior was predictable from motor variability using a statistically optimal model

For each unperturbed trial (constant K) in a given experimental condition, motor variability (noise) was computed as the SD (mm/s) of each striking velocity series ($N = 25$ strikes) measured during the two post-training phases. This was done for all three conditions. We analyzed these motor variability measures with a third LMM that included experimental condition as the within-participants fixed factor, expertise and class as the between-participants fixed factors, and the participant-specific intercept as a random factor [$\chi^2(1) = 17.3,\ p < .0001$]. The effect of expertise was significant [$F(2,72) = 12.3,\ p < .0001$; see Fig. 3.4a–b] and was not affected by condition, class, or by the condition $\times$ class interaction [$F \leq 1.39,\ p \geq .25$]. Post hoc contrasts between groups suggested that the percussionists performed better (lower motor noise) than the other two groups of participants [$|t(72)| \geq 4.23$, $p \leq .0002$] who performed similarly [$|t(72)| < 1$]. The effect of class was only marginally significant [$F(1,36) = 4.01,\ p = .053$], with slightly higher motor variability in the class of haptic-positive participants compared to the other class.

Importantly, the effect of condition was significant [$F(2,72) = 75.4,\ p < .0001$]: from auditory-only to haptic-only to bimodal condition, the amount of motor noise decreased monotonically [$|t(72)| \geq 3.18$, $p \leq .0065$ for all comparisons]. The effect of condition was also modulated by class [$F(2,72) = 7.62,\ p = .001$]. We were particularly interested in the effect of integrating the haptic feedback with the acoustical feedback on the achievement of the motor goal: minimization of the variability of the striking velocity. We took the pairwise differences in motor variability between the unimodal conditions and the bimodal condition. Independent two-sample t tests (two tailed) revealed that motor noise in the bimodal condition differed to a greater degree from motor noise in the auditory-only condition for the haptic-positive class than

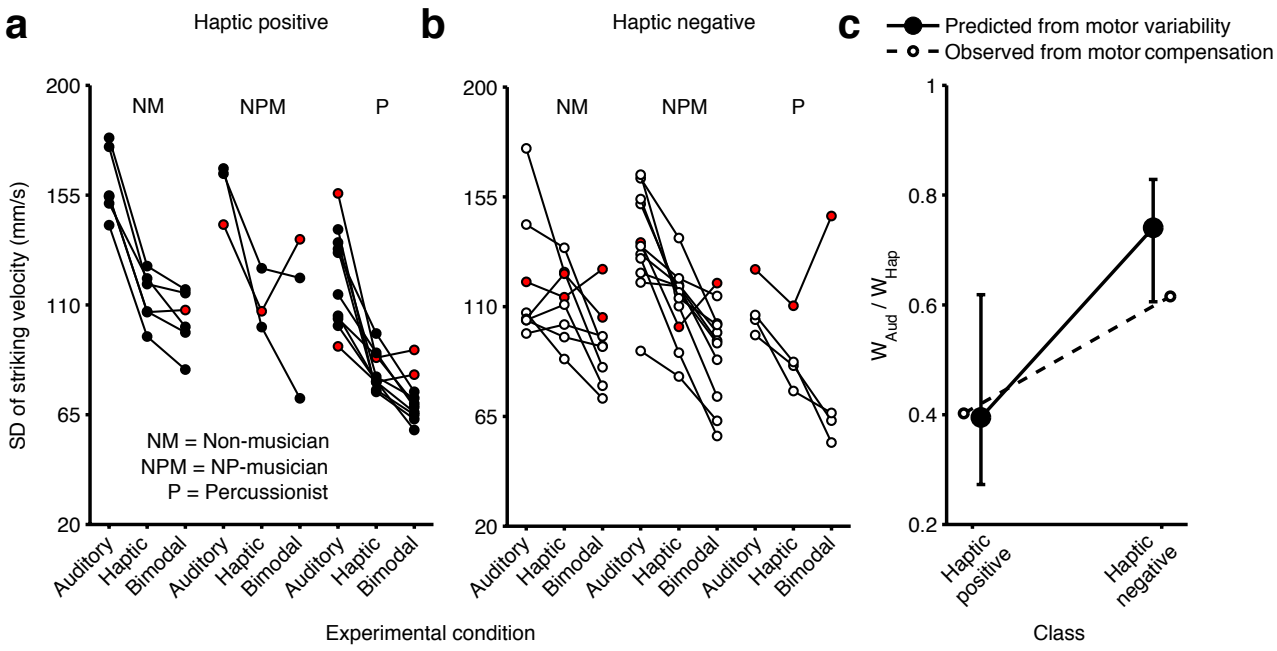

Fig. 3.4 Across-trial average of motor variability (SD, mm/s) as a function of experimental condition, measured in different expertise groups, and categorized by a factor of whether a participant had a positive (**a**) or negative (**b**) estimate of the fixed effect of K in the haptic-only condition. **c** Ratio of auditory weight to haptic weight (W_{Aud}/W_{Hap}) as predicted from the maximum-likelihood account of modality-specific motor variability in unperturbed trials, and as observed from the motor compensation in perturbed trials. Solid black dots show the participants' medians. Error bars indicate 95% confidence intervals (estimated by a 10,000-sample bootstrap).

for the other class [$t(40) = 2.15$, $p = .037$], whereas the differences in motor variability between the haptic-only condition and the bimodal condition were similar for the two classes [$t(40) = -1.42,\ p = .16$]. Qualitatively, these results indicated that the haptic feedback was a more reliable source of sensory information for the haptic-positive participants than for the haptic-negative participants. The higher reliability of the haptic feedback for the haptic-positive participants was reflected by its stronger uncertainty-reduction effect on the control signal for motor stability in the bimodal condition.

A quantitative account for the haptic-positive participants heavier reliance on the haptic

feedback in the audio-haptic context was achieved here. Assuming the principle of ML optimization, the weight (W) assigned to each control signal for striking velocity should be proportional to its own reliability (r) so as to minimize the uncertainty of an integrated signal. We subtracted the inverse of haptic-only motor variance (σ^2) and the inverse of auditory-only motor variance from the inverse of bimodal motor variance to estimate r_{Aud} and r_{Hap}, respectively.

Four haptic-positive participants and four haptic-negative participants did not strictly follow the ML-predicted pattern (Fig. 3.4ab, red markers): their motor noise in the bimodal condition was not strictly lower than both unimodal conditions, leading to negative estimates for r_{Aud} and/or r_{Hap} that are not computationally meaningful. In six out of the eight participants, however, their motor noise in the bimodal condition was lower than one of the two unimodal conditions, indicating the benefit from multisensory inputs during the control of striking velocity. An independent two-sample t test for the remaining individuals revealed significantly higher values of W_{Aud}/W_{Hap} ($= r_{Aud}/r_{Hap}$) in the 19 haptic-negative participants than in the 15 haptic-positive participants [$t(32) = -2.07$, $p = .023$, one-tailed]. This result is in line with the more pronounced haptic dominance as postulated from the asymmetrical effects of crossmodal perturbation congruency across audition and haptics (Fig. 3.4c). We also noticed marked inter-individual differences in the ML-predicted sensory weighting and overestimated values of W_{Aud}/W_{Hap} for the haptic-negative participants in particular. This result raises a concern (see Chapter 4: Discussion): the quality of proprioceptive feedback resulting from wrist movements might not be strictly equivalent across the experimental conditions with and without the haptic surface, which was not present in the auditory-only condition.

Chapter 4

Discussion

We investigated compensatory motor effects of multisensory processes during percussive sonic interactions with objects of unpredictable surface stiffness. Prediction errors concerning the striking kinematics (wrist velocity) were imposed by altering the impact-related haptic and auditory feedback (alone or simultaneously) while participants repeatedly struck a target virtual sounding object. We uncovered evidence for a dominance of haptics over audition in the control of striking kinematics. The observed dominance effects appeared to be in agreement with a maximum-likelihood (ML) prediction based on modality-specific motor noise. Each sensory modality was weighted as a function of its relative reliability so as to reduce the uncertainty of control signal for motor stability. Developing such a statistically optimal strategy seems not to be limited to purely perceptual tasks (Ernst & Banks, 2002; Gepshtein & Banks, 2003), or motor tasks involving movements without goal-directed object contact (Sober & Sabes, 2003; Saunders & Knill, 2004; Ronsse et al., 2009; Frissen et al., 2011).

Given the nature of our experimental manipulation of the haptic feedback, an unexpected increase in the object stiffness would result in a sudden decrease in the perceived surface deformation by the stylus, but also a higher magnitude and faster increment rate of the resistive

force generated during the stylus-object contact. Striking an object with higher velocity (greater energy supplied to the vibrating system) would instead always generate an increase in both the deformation and the force cues (Fig. 2.2c–f).

These conditions essentially provide a testing ground for whether the deformation and force cues were preferentially attended to by different participants faced with the task of estimating compensatory motor variables for subsequent actions. A large inter-individual variability emerged in the haptically induced motor responses, which could be most parsimoniously explained in terms of a selective reliance on different tactual error signals for the motor control. Fifty-five percent of the participants (haptic negative) likely tended to compensate for the altered force cues by striking more softly (lower striking velocity) on a stiffer object, whereas another 45% (haptic positive) likely tended to compensate for the altered surface deformation by striking harder on the stiffer object in order to indent its surface as deeply as had been experienced prior to the perturbation. Previous evidence has shown that the two primary types of tactual cues of surface stiffness during active explorations of deformable objects (LaMotte, 2000; Friedman et al., 2008) undergo complicated interactions when subserving a purely perceptual discrimination of compliance (see Di Luca & Ernst, 2014 on a Bayesian decision theory of stiffness perception; and Tan, Durlach, Beauregard & Srinivasan, 1995 for a proposal that the force integral over deformation serves as a potential cue). Our study using an active kinematics-control task instead indicates that participants solve the sensorimotor control problem by exploiting the sensory evidence that potentially best signals a fast motor estimate (not necessarily being absent from bias), which minimizes the expected loss (Körding & Wolpert, 2006) as a consequence of deviation from the targeted striking kinematics. The exact origin of the observed cue selectivity for motor control remains an open question (although a working hypothesis is proposed below). Further experimentation is necessary to dissociate the two types of tactual cues, which is challenging for a model-based haptic device such as ours.

Notably, we found that 10 out of 14 percussionists were haptic positive; eight of these 10 reported being mallet percussionists (most of them specialized in xylophone, marimba, etc.). Increased sensitivity to a particular type of error signal during percussive events could be shaped by many years of experience in striking objects of varying degrees of stiffness with a diversity of mallets that also vary in stiffness. To stabilize a manipulative behavior on objects, most people potentially employ a mechanism (an internal model of the object properties, Flanagan & Wing, 1997) that seems not to require special expertise: grasp stability is achieved by adjusting the grip force in anticipation of the changes in resistive force generated by a tapped object (LaMotte, 2000). Percussionists, however, have potentially developed a different strategy. They routinely relax the grip so as to allow the hand-held mallet to rotate freely around a fulcrum point (commonly between the index finger and thumb) in a vertical plane (Dahl, 2006). They in fact make an effort to keep the wrist joint as compliant as possible (to prevent fatigue) during repeated percussive gestures, which is achieved through unique physiologically efficient reciprocal activities of the wrist muscles (Fujii et al., 2009). A habitual strategy of monitoring wrist-rotation trajectories (position-related tactual signals encoded via joint receptors and muscle afferents) rather than the grip force would possibly facilitate a percussionist's sensitivity to the nuances in the magnitude of object indentation by the stylus.

Previous psychophysical evidence has shown putative dominance of haptic information over auditory information in perceptual tasks of estimating object textures (Lederman, 1979; Lederman et al., 2002), judging apparent motions (Soto-Faraco, Spence & Kingstone, 2004a) or spatial localizations (Caclin, Soto-Faraco, Kingstone & Spence, 2002) or of identifying materials (Giordano et al., 2012) within audio-haptic situations. Our results provide the quantitative evidence of haptic dominance over audition in a complex sonic-interaction context. The congruency effect across audition and haptics was asymmetrical, with the acoustically induced motor responses being modulated more strongly by a synchronized haptic perturbation than the

haptically induced motor responses being modulated by a synchronized acoustical perturbation. The stronger weighting of haptics compared to audition might, at least in part, be attributable to a bias towards focusing on haptics that is more tightly linked to the exploration and exploitation of effector-related information during active sensorimotor activities. This superior modality-specific utility could be learned through long-term tactual experience of object manipulations/explorations. For instance, the identification of aggregate walked-upon materials (e.g., gravels of different sizes) through locomotor movements appeared to be dominated by the kinesthetic information that would have most promptly signaled a potential postural instability (Giordano et al., 2012). We also noticed that the acoustical perturbation did lead to a small but observable effect on the haptically induced motor responses, drawing a parallel with analogous reciprocal audio-tactile interactions (albeit with an asymmetry in favor of touch) at perceptual levels of motion processing (Soto-Faraco et al., 2004a). This result contrasts with the highly imbalanced modality preference during speech sensorimotor control (Lametti et al., 2012), i.e., motor compensation for somatosensory perturbation (altered motion path of the jaw) was not changed by simultaneously applied auditory perturbation (altered frequency). It is unknown whether this disagreement is due to a neurophysiological distinction between the two sensorimotor systems (limb- vs. vocal-specific) or is the result of a methodological incompatibility: we employed an ecologically relevant manipulation of crossmodal prediction errors based on audio-haptic stiffness, whereas Lametti et al. (2012) manipulated the uncorrelated cues of movement path of the jaw and formant frequencies of the utterance.

The haptic-positive participants' heavier reliance on the haptic feedback in the audio-haptic context is in fact consistent with the ML prediction based on modality-specific motor noise—sensory weighting relies on reliability. When the goal of repeated movements is to converge the kinematic outputs toward a targeted stationary state, the reliability of a sensory feedback relevant to the action can be reflected by the extent to which motor variability is

minimized in the presence of that source of sensory feedback. According to our results, however, the ideal ML schema appears to some extent to overestimate the relative weight of auditory feedback (Fig. 3.4c). A potential interpretation holds that the reliability of proprioceptive inputs resulting from wrist movements was not strictly equivalent across the experimental conditions with and without the existence of a haptic surface in the haptic-only and bimodal conditions on the one hand, and in the auditory-only condition on the other. It is likely that people decreased the reliance on proprioception during the haptic-only and the bimodal trials because of the impact-generated vibrations transmitted through the stylus to the muscle spindles and joint receptors near the wrist, causing proprioception to be distorted. If so, the reliability of the touch-related control signal, which has been computed ideally as the difference between the inverse of bimodal motor variance and the inverse of auditory-only motor variance, appears to be underestimated by the simple subtraction. Note, however, that such a concern would not invalidate our primary ML predictions regarding either the asymmetrical interactions between haptically and acoustically induced motor compensations (as predicted, ratios of auditory to haptic weight were consistently less than 1 across participants) or the inter-class difference in sensory weighting (Fig. 3.4c).

If we assume that the haptic-positive participants were sensitive to the altered object deformation, an alternative account for their greater disregard for the altered auditory feedback in the bimodal condition holds that the ML rule breaks down if a process of causal inference judges that the different sensory signals originate from weakly correlated sources (Körding, Beierholm, Ma, Quartz, Tenenbaum & Shams, 2007; Parise, Spence & Ernst, 2012). Motor adaptation also depends on the relevance of the sensory errors such that a probabilistic sensorimotor model is structured in the central nervous system to underweight irrelevant sensory errors (Wei & Körding, 2009). Long-term experience implies that striking both rigid and deformable objects is naturally accompanied by impact sounds and resistive force. Short-term, repeated manipulations of the

haptic device may have additionally given rise to people's awareness that the impact sounds fail to inform the representation of the object's deformation, although they still convey spatial information about the contacts with the surface (DiFranco, Beauregard & Srinivasan, 1997). The haptic deformation cues are thus less causally correlated with the auditory cues than are the force (or vibrotactile) cues. The auditory cues were thus more likely to be ignored if a participant showed predominantly greater sensitivity to the deformation-related prediction errors.

Control signals for motor output are often characterized by temporal lags (e.g., feedback- vs. feedforward-driven motor commands). Computationally attractive models (e.g., Kalman-based filtering) have been proposed for understanding how (delayed) sensory and feedforward information are averaged optimally over time to drive efficient state estimation (Goodwin & Sin, 1985; Wolpert et al., 1995; Deneve et al., 2007). By what mechanisms then are the control signals based on different sensory feedback (obtained through contact with external object) combined to guide a stabilized pattern of striking kinematics? The present study indicates that a minimum-uncertainty strategy underlying the intersensory integration for the perception of object stiffness (Di Luca & Ernst, 2014) is also likely adopted by the sensorimotor system to plan the wrist kinematics for reliable sonic interactions with an object. Although percussion expertise is likely, to some extent, to shape the sensitivity to a particular type of haptic error that drives motor compensation, the observed reliability-based sensory weighting for motor control within each class of individuals turns out to operate independently of musical experience (note that there were indeed 32% of non-percussionists being haptic positive and 29% of percussionists being haptic negative). Musicianship did not affect the cross-condition pattern of motor variability that occurred in each class of individuals (Fig. 3.4a–b). The inter-class difference in sensory weighting was actually explained by a substantially stronger motor-noise-minimization effect of the haptic feedback (a comparison between the auditory-only and bimodal conditions) for the haptic-positive class (Fig. 3.4a) than for the other class (Fig. 3.4b). This finding echoes the

coherent ML-based predictions of sensory fusion across different learning stages for motor execution (which indicates the absence of expertise effect) in a previous study that employed a bimanual coordination task involving cyclical wrist movements (Ronsse et al., 2009).

Chapter 5

Future directions

5.1 Efferent and afferent control-signal components

One possible avenue for further research concerns whether the efferent (central motor commands) and the afferent (sensorimotor transformation based on auditory, proprioceptive, and tactile feedback) signal components for the control of wrist kinematics can be distinguished and quantified. Aiming at quantifying the weighting of the impact-generated haptic and acoustical feedback, the present study has in fact treated the central motor commands and the proprioceptive feedback together as a "compound" signal, which could be computationally cancelled out when contrasting the bimodal condition with each unimodal condition. A behavioral approach to separate the feedforward and sensory components could involve systematically supplementing the current experimental design (auditory-only, haptic-only and bimodal) with two other control conditions: measuring motor variability 1) in trials with neither the impact-related haptic nor acoustical feedback (only feedforward and proprioceptive signals), and 2) in trials with the impact-related acoustical feedback only (no haptic surface) but using techniques in which the quality of the proprioceptive feedback could be degraded (thus only feedforward and auditorily

transformed components). The proprioceptive feedback about the wrist movements could be manipulated via non-invasive approaches, e.g., by applying coupled agonist-antagonist vibratory stimulations at the same frequency (thus without biasing overt motor outputs, Gilhodes, Roll & Tardy-Gervet, 1986) upon the dorsal and palmar portions of the wrist's tendons (Bock, Pipereit & Mierau, 2007; Ronsse et al., 2009). Thus the problem of five unknown control-signal components, i.e., relevant to the feedforward motor command, the impact-related acoustical and haptic feedback, the proprioception in the trials with and without the presence of the haptic surface, could have unique solutions by considering the motor variabilities measured in the five experimental conditions. An additional benefit of quantifying the potential change in the reliability of proprioception across the haptic and non-haptic conditions is that we might by this means confirm the hypothesis about the observed slightly overestimated weights of audition (see Chapter 4: Discussion).

5.2 Dissociation of deformation errors from force feedback

The results reported in this study suggest a selective reliance on different tactual error signals (deformation vs. force cues) for wrist velocity control among different participants. Many probabilistic models of motor behavior postulate that a movement plan is chosen so as to optimize a certain "measure of accuracy" (i.e., loss function defined on task-relevant errors, from a statistical learning perspective, Wolpert, 2007; Wolpert & Landy, 2012). In particular, the error measure selected to be penalized appears to be contextually dependent, with small errors being penalized quadratically (minimizing average squared error), and large errors being penalized almost linearly (Körding & Wolpert, 2004b). To further examine how availabilities of different tactual signals influence motor decisions, another research avenue would experimentally dissociate the deformation cues from the force cues. One might consider ERGOS, a high-fidelity

haptic device capable of rendering rigid objects using a maximum efficient force of 200 N (stiffness $\geq$ 40,000 N/m) per degree of freedom (Florens, Luciani, Cadoz & Castagné, 2004); alternatively, one might consider an admittance-based device (Van der Linde, Lammertse, Frederiksen & Ruiter, 2002), which is controlled to move proportionally to the user-exerted force, whereby the penetration depth of the virtual object could be minimized to the lowest possible extent. Another related avenue for further research could examine the extent to which the observed cue selectivity is shaped through specificity of mallet-based percussion training. Follow-up studies could include "non-mallet" percussionists who have been trained most of the time to play with their bare hands and fingers (e.g., tabla and conga players). Issues such as certain strict criteria for participant selection might be considered, because modern percussionists are usually known for their versatility across a broad range of percussion instruments.

Appendix A

Synthesis models for stimuli

The physical modeling of impact sound was based on a compliant collision model (impactor-object interaction) with nonlinear damping (Hunt & Crossley, 1975; Marhefka & Orin, 1999). The time-varying resistive force $f(t)$ generated during a collision is modeled as:

$$f(t) = -kx(t)^{\alpha} - \lambda x(t)^{\alpha} v(t), x(t) \geq 0 \tag{A.1}$$

where $v(t)$ represents the time-varying compression velocity (m/s) of the impactor, and $x(t)$ is the time-varying displacement (m) of the impactor relative to the surface of the object; k represents the stiffness coefficient, corresponding to the acoustical K throughout the main texts; α is a geometry-dependent exponent, which was set to 1.5, so as to model a sphere-plate contact subject to the Hertz's law (Landau & Lifshitz, 1981). The parameter λ is a damping weight (Marhefka & Orin, 1999). Due to the dissipative nature of the model, changing $v(t)$ could modify the amplitude and spectral properties of the synthesized impact sounds even under circumstances in which the acoustical K is kept constant. Specifically, Avanzini & Rocchesso (2004) derived an analytical equation that relates the model-simulated contact time τ to the physical parameters of

the collision model (and consequently the acoustical properties of the impact sounds):

$$\tau = \left(\frac{m}{k}\right)^{\frac{1}{1+\alpha}} \cdot \left(\frac{\mu^2}{1+\alpha}\right)^{\frac{\alpha}{1+\alpha}} \cdot \int_{v_{in}}^{v_{out}} \frac{1}{(1+\mu v)\cdot\left[-\mu(v-v_{in}) + \log\left|\frac{1+\mu v}{1+\mu v_{in}}\right|\right]^{\frac{\alpha}{1+\alpha}}} \tag{A.2}$$

where m is the impactor mass; the remaining parameters are part of Equation (A.1), and $\mu = \lambda/k$ is a mathematically convenient term representing the material's viscoelastic characteristic (Marhefka & Orin, 1999). Equation (A.2) states that the contact time τ depends only on μ, the exponent α and the ratio m/k, in addition to the initial velocity at impact, v_{in}. Because neither m nor k affects the value of the integral, it follows that for a given value of v_{in}, the dependence $\tau \sim (m/k)^{1/(1+\alpha)}$ always holds.

Sound stimuli were synthesized in real time by varying the parameters k and v_{in}. All other synthesis parameters were fixed: the impactor mass m was set to 0.5 kg; the damping weight λ was set to 0.3 $\mathrm{gm}^{-1.5}\mathrm{s}^{-1}$; and finally, an internal friction coefficient $\tan\phi$ was set to 36.8 (arbitrary units). The coefficient $\tan\phi$ was developed with the intent of extracting an acoustical measure that uniquely specifies the material of an object (Wildes & Richards, 1988). More specifically, $\tan\phi$ is expressed as

$$\tan\phi = \frac{\alpha_i}{\pi f_i} \tag{A.3}$$

where f_i is the frequency of ith vibration mode, and α_i is a decay-related quantity, defined as the reciprocal of the time t_e required for the amplitude to decay to $1/e$ of its starting value. Damping rate increases with internal friction, so that highly damped materials, such as rubber, are characterized by higher $\tan\phi$ values. Here the fundamental vibrational mode was set to be characterized by a t_e of approximately 87 ms.

Based on this analytical property, a relation between the contact time and the time-varying spectral centroid of the impact sound was quantified by characterizing a close mapping between

the physical parameters of the impact force and the hardness-related auditory cues (Avanzini & Rocchesso, 2004). Using an absolute magnitude-estimation procedure, Avanzini & Crosato (2006) discovered that changing k significantly biased people's subjective hardness ratings of tapped virtual objects, although the objective stiffness was in fact kept constant.

The force-rendering algorithm in the haptic collision model was a simplification of Equation (A.1). The geometry-dependent exponential coefficient α was set to 1 for haptic simulations. Similar to the audio rendering, a dissipative component was added to the synthesized resistive force, and the damping weight λ was set to 0.5 (any value from 0 to 1 was acceptable in the Phantom haptic system).

Appendix B

Spectrotemporal analyses of impact sounds

B.1 Extraction of acoustical features (descriptors)

We investigated the relationship between the synthesis parameters (K and striking velocity at impact) of the physical modeling and the acoustical features of the synthesized impact sounds.

The impact sounds were synthesized using the struck-bar model, which was the one used for generating the experimental stimuli (see Chapter 2: Methods). A total of 400 sounds was investigated. The synthesis parameter acoustical K varied across 20 log-spaced levels ranging from 10^2 to 10^5 N/m$^{1.5}$. For each of the 20 K levels, 20 different striking velocities (another sound synthesis parameter) were chosen from a wide range of values (see Table B.1). Acoustical features were extracted with previous methods (Giordano & McAdams, 2006; Giordano et al., 2010; McAdams et al., 2010) in auditory psychomechanics. Two different representations were used for extracting a total of 12 features (descriptors): 1) the amplitude envelope, and 2) a simulation of the signal transformations taking place in the human peripheral auditory system.

The amplitude envelope $E(t)$ of an impact sound with time-varying amplitude $x(t)$ can be

Table B.1 Acoustical K and striking velocities used to synthesize impact sounds for acoustical analysis.

	Striking Velocity		
Acoustical K	Min.	Mean	Max.
2.000	158	427	891
2.157	172	524	944
2.315	321	532	818
2.473	227	489	847
2.631	257	558	819
2.789	158	541	929
2.947	235	608	974
3.105	203	642	1058
3.263	117	562	1010
3.421	280	621	1134
3.578	157	521	950
3.736	255	707	1164
3.894	110	528	818
4.052	290	599	1041
4.210	48	672	1163
4.368	158	667	1118
4.526	58	693	1103
4.684	271	599	903
4.842	49	520	1013
5.000	141	429	848

Notes. Acoustical K varied across 20 log-spaced levels ranging from 10^2 *to* 10^5 *(N/m*$^{1.5}$*). For each level of K, 20 different striking velocities (another sound synthesis parameter) were chosen from a wide range of values.*

represented as:

$$E(t) = |x(t) + i\mathrm{H}[x(t)]| \tag{B.1}$$

where H[·] is the Hilbert transform (Hartmann, 1997). Hearing-range amplitude fluctuations were attenuated by forward-reverse filtering $E(t)$ using a third-order zero-phase-distortion Butterworth filter with a low-pass cutoff frequency of 16 Hz. Three acoustical descriptors (α, $ED_{3\ \mathrm{dB}}$, $ED_{10\ \mathrm{dB}}$) were extracted based on $E(t)$ converted to dB relative to the peak digital amplitude of 1. Specifically, α measures the slope of the least-squares line fitted to a 30-ms envelope portion

starting from 5 ms after the peak level; $ED_{3\text{ dB}}$ and $ED_{10\text{ dB}}$ characterize the duration in ms over which the amplitude envelope exceeds the -3 and -10 dB threshold with respect to the peak level, respectively (ED refers to equivalent duration).

Another set of descriptors was extracted based on a simulation of the signal processing that takes place in the peripheral auditory system (McAdams et al., 2004). The stimulation produced the time–varying power at the output of a set of cochlear filters. The center frequencies of the cochlear filters f_c (range: 30-16,000 Hz) were uniformly spaced on a frequency scale derived from measures of masked detection thresholds in normal–hearing humans: the Equivalent Rectangular Bandwidth (ERB)–rate scale (Moore & Glasberg, 1983) reflects the spacing of auditory filters along the basilar membrane. An important descriptor extracted from this representation was $\tan\phi_{\text{aud}}$, a unitless measure of the damping of vibration in struck solids (Wildes & Richards, 1988) associated with human judgments of sounding-object materials (Giordano & McAdams, 2006; Giordano et al., 2010). Intuitively, the higher $\tan\phi_{\text{aud}}$, the faster the relative energy decay in cochlear filter channels with higher center frequencies. More specifically, $\tan\phi_{\text{aud}}$ is expressed as

$$\tan\phi_{\text{aud}} = \frac{\sum_i \frac{\alpha_i}{\pi f_{ci}} p_i}{\sum_i p_i} \tag{B.2}$$

where α_i is the damping factor for the power output at the ith cochlear filter, as estimated with linear regression methods. The damping factor α_i for the signal output from each channel is computed using the regression model $\log_e(p_i) = a + bT$, where p_i is power, T is time, and the relation between the regression slope b and the damping factor α_i is defined as $b = -\alpha_i/2$. The regression model was applied to the signal from peak power to a threshold level of the $1/e$ of peak power. f_{ci} is the center frequency in Hertz of the ith cochlear filter, and p_i is the total power

at the output of the *i*th cochlear filter.

The temporal resolution of the output at each cochlear filter was subsequently decreased to 100 Hz, where the 10-ms resolution corresponds roughly to the auditory window of temporal integration in humans (Plack & Moore, 1990). In a second step, the down-sampled energy output from the cochlear filters was raised to the power of 0.25 to yield an approximate measure of the time-varying loudness within each cochlear filter channel (Hartmann, 1997, p. 66), termed "specific loudness" (Zwicker & Fastl, 1999).

Finally, the time-varying loudness and spectral center of gravity (SCG) were extracted as the sum of the specific loudnesses and as the specific-loudness-weighted average frequency on the ERB-rate scale, respectively. SCG captures the auditory attribute of brightness (Grey & Gordon, 1978; McAdams, Winsberg, Donnadieu, De Soete & Krimphoff, 1995). The measurement unit for loudness was pseudo–sones (p.s.), independent of the actual presentation levels. Four loudness-related descriptors were derived: the attack value, Lou_{att}, which is the average loudness of the first 10-ms portion of the signal; the mean value over the whole duration, Lou_{mea}; and the slope of the initial (30 ms starting from the peak loudness) and final (the last 80 ms of the loudness function) portions of the temporal function of loudness, Lou_{sl1} and Lou_{sl2}, respectively. Three SCG–related descriptors were extracted: the attack value SCG_{att}, which is the SCG of the first 10 ms of the signal; the mean value over the whole duration, SCG_{mea}; and the slope of the initial portion of the temporal function of SCG (i.e., the first 30 ms of the SCG signal starting from the peak value), SCG_{slo}. The slope measures were extracted by means of linear regression over a portion of the temporal function. Finally, the effective duration of the signal, Dur, was defined as the temporal extent over which the loudness exceeded a fixed threshold of 1.0 p.s. (i.e., a level sufficiently higher than the background noise).

The set of 12 acoustical descriptors (α, $ED_{3\ \mathrm{dB}}$, $ED_{10\ \mathrm{dB}}$; $\tan\phi_{\mathrm{aud}}$, Lou_{att}, Lou_{mea}, Lou_{sl1} and Lou_{sl2}; SCG_{att}, SCG_{mea}, SCG_{slo}; Dur) was computed for each of the 400 impact sounds.

Table B.2 Spearman rank correlation coefficients between acoustical features of impact sounds. df = 398.

	Dur	$\tan\phi_{\text{aud}}$	Lou_{att}	Lou_{mea}	Lou_{sl1}	Lou_{sl2}	SCG_{att}	SCG_{mea}	SCG_{slo}	α	$ED_{3\text{ dB}}$	$ED_{10\text{ dB}}$
Dur	–	-0.07	0.77^{***}	0.87^{***}	-0.73^{***}	-0.87^{***}	0.53^{**}	-0.72^{***}	-0.37^{**}	-0.53^{**}	-0.53^{**}	-0.54^{**}
$\tan\phi_{\text{aud}}$		–	-0.57^{***}	-0.45^{**}	0.59^{***}	0.08	-0.72^{***}	-0.45^{***}	0.64^{***}	0.72^{***}	0.71^{***}	0.69^{***}
Lou_{att}			–	0.98^{***}	-0.99^{***}	-0.65^{***}	0.94^{***}	-0.16^{*}	-0.71^{***}	-0.94^{***}	-0.94^{***}	-0.94^{***}
Lou_{mea}				–	-0.96^{***}	-0.76^{***}	0.85^{***}	-0.33^{***}	-0.62^{***}	-0.86^{***}	-0.86^{***}	-0.86^{***}
Lou_{sl1}					–	0.62^{***}	-0.95^{***}	0.12^{*}	0.73^{***}	0.95^{***}	0.95^{***}	0.95^{***}
Lou_{sl2}						–	-0.43^{**}	0.68^{***}	0.36^{**}	0.43^{**}	0.43^{**}	0.43^{**}
SCG_{att}							–	0.15^{*}	-0.75^{***}	-0.98^{***}	-0.99^{***}	-0.99^{***}
SCG_{mea}								–	-0.01	-0.14^{*}	-0.13^{*}	-0.13^{*}
SCG_{slo}									–	0.75^{***}	0.76^{***}	0.76^{***}
$ED_{3\text{ dB}}$											–	0.99^{***}
$ED_{10\text{ dB}}$												–

$^{*}p < .01$; $^{**}p < .001$; $^{***}p < .0001$

Notes. Dur = duration; Lou = loudness; SCG = Spectral Center of Gravity; att = attack; mea = mean; sl1 = initial slope; sl2 = tail slope; slo = slope; p.s. = pseudo–sones; ERBr = Equivalent Rectangular Bandwidth rate; ED = Effective Duration.

B.2 Inter-descriptor correlation and clustering analysis

We computed the pairwise Spearman rank correlation coefficients ρ between the 12 acoustical descriptors (Table B.2; all $df = 398$). Some acoustical descriptors were highly correlated, e.g., the two attack properties SCG_{att} and Lou_{att} ($\rho = .94$). Another highly correlated pair of descriptors, among many others, was the Lou_{sl2} and Dur ($\rho = -.87$), indicating that the rate of loudness decay in the tail of the signal and the signal's effective duration both specify the damping-related characteristics of the impact sound. Note also that the attack-property specifier SCG_{att} (Giordano et al., 2010) was only weakly correlated with the two damping-related features Lou_{sl2} and Dur ($|\rho| \leq .53$).

We took a clustering analysis to reduce groups of correlated descriptors to a single variable, following the methodology presented in a previous study by Giordano et al. (2010). The choice of the number of clusters relies on the concept of degenerate components, i.e., class variables (CLs) that explain less variance than one single input variable. The number of feature clusters in this algorithm corresponds to the first clustering level where the number of non-degenerate CLs for each cluster equals one, and where the variance of the input data explained by the final class variable decomposition is greater than or equal to the variance explained by a dimensionality reduction of all of the input data minus degenerate CLs.

The clustering procedure revealed three independent classes (average inter-class $|\rho| = .41$, $SD = .28$), with each one capturing the shared aspects of a group of highly correlated descriptors (Fig. B.1a). The first class CL_1 contained a single feature $\tan\phi_{\mathrm{aud}}$. The second class CL_2 included eight acoustical descriptors $\{\mathrm{Lou}_{att}, \mathrm{Lou}_{mea}, \mathrm{Lou}_{sl1}, \mathrm{SCG}_{att}, \mathrm{SCG}_{slo}, \alpha, ED_{3\ \mathrm{dB}}, ED_{10\ \mathrm{dB}}\}$. Notably, the attack descriptors, SCG_{att} and Lou_{att}, together with two other loudness descriptors, Lou_{mea} and Lou_{sl1}, were accurate acoustical specifiers for the properties of the mallet/plate interactions (Giordano et al., 2010). The third class CL_3 including $\{\mathrm{Lou}_{sl2}, \mathrm{SCG}_{mea},$

Table B.3 Spearman rank correlations (ρ) between acoustical features and three reduced class (CL) variables according to clustering analysis.

Feature	Class	ρ
$\tan\phi_{\text{aud}}$	CL_1	1.0
Lou_{att}	CL_2	0.98
Lou_{mea}	CL_2	0.92
Lou_{sl1}	CL_2	−0.98
SCG_{att}	CL_2	0.98
SCG_{slo}	CL_2	−0.79
α	CL_2	−0.99
$ED_{3\text{ dB}}$	CL_2	−0.98
$ED_{10\text{ dB}}$	CL_2	−0.99
Dur	CL_3	0.95
Lou_{sl2}	CL_3	−0.93
SCG_{mea}	CL_3	−0.87

Notes. Abbreviations of acoustical features are as in Table B.2. See text for details.

Dur}, appeared to be related to the acoustical damping characteristics.

The total variance explained by the non-degenerate components was 92%. Overall, the three classes accounted well for the original acoustical descriptors: high correlations were found between the acoustical descriptors and their respective class, grand average $|\rho_s| = .95$, $SD = .063$ (see Table B.3). The independence of $\tan\phi_{\text{aud}}$ (CL_1) against the other acoustical features was not surprising, due to the fact that the parameter $\tan\phi$ (an internal friction, measuring damping rate of material, Wildes & Richards, 1988; also see Appendix A) was fixed in the synthesis model.

B.3 Acoustical correlates of stiffness coefficient and impact velocity

We aimed to examine the relationships between the 12 acoustical features and the two synthesis parameters acoustical K and impact velocity (v). We randomly took 20 impact sounds that were generated using the 20 different log-spaced K values (10^2 to 10^5 N/m$^{1.5}$). Given the results of the clustering analysis, we first reversed the rank order of several acoustical descriptors,

by multiplying them by -1, to ensure that the descriptors within each CL always covaried positively. Then, the descriptors belonging to the same CL were reduced to a single principle component (PC) using the robust PCA approach (Hubert, Rousseeuw & Vanden Branden, 2005). We computed the Spearman rank correlations between each PC and the two synthesis parameters, respectively. The above procedures were repeated 1000 times, with each time choosing randomly one of the 20 possible striking velocities for each K level, resulting in six distributions (1000 values in each) of correlation coefficients (Fig. B.1b). The PC related to $\tan\phi_{\text{aud}}$ was highly correlated with acoustical K, the average (95% CI) rank correlation coefficient was $-0.72\ (-0.91, -0.51)$, but was weakly correlated with the impact velocity, the average (95% CI) rank correlation coefficient was $0.11(-0.24, 0.44)$. The other PCs were correlated with K and impact velocity with the same sign $(-)$ of correlation coefficients, indicating that the acoustical features (except $\tan\phi_{\text{aud}}$) were influenced by K (intrinsic property of the object) and striking velocity (extrinsic kinematic property determined by a user) in the same direction.

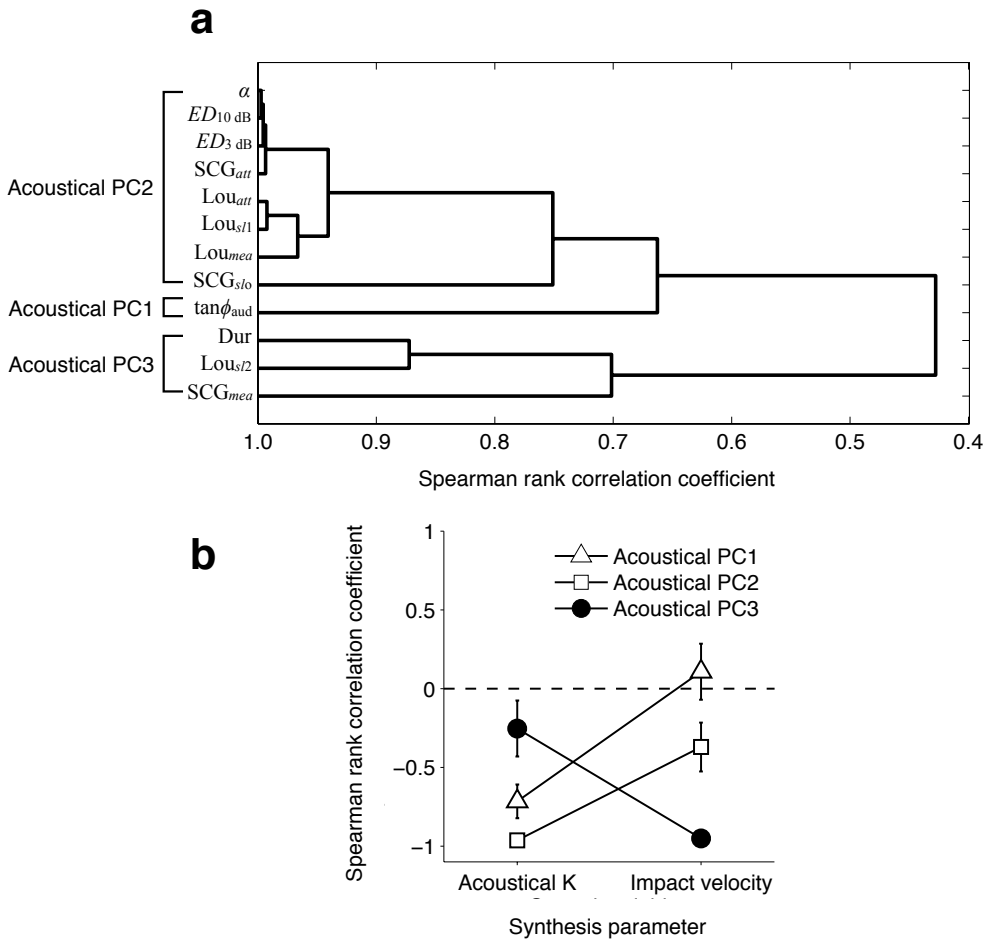

Fig. B.1 Acoustical correlates of stiffness coefficient and striking velocity in synthesized impact sounds. **a** Clustering analysis of acoustical descriptors. **b** Mean Spearman rank correlation coefficients ($df = 18$) between each principle component (PC) and two synthesis parameters (acoustical K and impact velocity). Error bars indicate $\pm$ SD ($N = 1000$).

Bibliography

Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Angel, R. W. (1976). Efference copy in the control of movement. *Neurology*, *26*(12), 1164–1164.

Avanzini, F. & Crosato, P. (2006). Integrating physically based sound models in a multimodal rendering architecture. *Computer Animation and Virtual Worlds*, *17*(3-4), 411–419.

Avanzini, F., Rath, M., Rocchesso, D., & Ottaviani, L. (2003). Low-level models: Resonators, interactions, surface textures. In D. Rocchesso & F. Fontana (Eds.), *The Sounding Object* (pp. 137–172). Firenze, Italy: Mondo Estremo.

Avanzini, F. & Rocchesso, D. (2004). Physical modeling of impacts: Theory and experiments on contact time and spectral centroid. In *Proceedings of the Conference on Sound and Music Computing*, (pp. 287–293)., Paris, France.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391–1397.

Berniker, M. & Körding, K. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, *11*(12), 1454–1461.

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *NeuroReport*, *12*(9), 1879–1884.

Bock, O., Pipereit, K., & Mierau, A. (2007). A method to reversibly degrade proprioceptive feedback in research on human motor control. *Journal of Neuroscience Methods*, *160*(2), 246–250.

Bresciani, J.-P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile tap perception. *Experimental Brain Research*, *162*(2), 172–180.

Cabe, P. A. & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 313–324.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile "capture" of audition. *Perception & Psychophysics*, *64*(4), 616–630.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science*, *9*(3), 211–214.

Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS One*, *5*(8), e12240.

Chaigne, A. & Doutaut, V. (1997). Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars. *Journal of the Acoustical Society of America*, *101*(1), 539–557.

Churchland, M. M., Afshar, A., & Shenoy, K. V. (2006). A central source of movement variability. *Neuron*, *52*(6), 1085–1096.

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, *32*(3), 1056–1060.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, *4*(1), 102–118.

Colgate, J. E. & Brown, J. M. (1994). Factors affecting the z-width of a haptic display. In *Proceedings of the IEEE 1994 International Conference on Robotics and Automation*, (pp. 3205–3210)., San Diego, CA.

Dahl, S. (2006). Movements and analysis of drumming. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music motor control and the brain* (pp. 125–138). Oxford, England: Oxford University Press.

Deneve, S., Duhamel, J.-R., & Pouget, A. (2007). Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of kalman filters. *Journal of Neuroscience*, *27*(21), 5744–5756.

Di Luca, M. & Ernst, M. O. (2014). Computational aspects of softness perception. In M. Di Luca (Ed.), *Multisensory softness–Perceived compliance from multiple sources of information* (pp. 85–106). London, UK: Springer–Verlag.

Diedrichsen, J., Verstynen, T., Lehman, S. L., & Ivry, R. B. (2005). Cerebellar involvement in anticipating the consequences of self-produced actions during bimanual movements. *Journal of Neurophysiology*, *93*(2), 801–812.

DiFranco, D. E., Beauregard, G. L., & Srinivasan, M. A. (1997). The effect of auditory cues on the haptic perception of stiffness in virtual environments. In *Proceedings of the ASME Dynamic Systems and Control Division*, volume 61, (pp. 17–22). American Society of Mechanical Engineers.

Driver, J. & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 217–220). Oxford, UK: Oxford University Press.

Ernst, M. O. & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*(1), 69–73.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303.

Ferris, D. P., Louie, M., & Farley, C. T. (1998). Running in the real world: Adjusting leg stiffness for different surfaces. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1400), 989–994.

Flanagan, J. R. & Wing, A. M. (1997). The role of internal models in motion planning and control: Evidence from grip force adjustments during movements of hand-held loads. *Journal of Neuroscience*, *17*(4), 1519–1528.

Fletcher, N. H. & Rossing, T. D. (1991). *The physics of musical instruments*. New York, NY: Springer–Verlag.

Florens, J.-L., Luciani, A., Cadoz, C., & Castagné, N. (2004). ERGOS: Multi-degrees of freedom and versatile force-feedback panoply. In *Proceedings of EuroHaptics*, (pp. 356–360)., Munich, Germany.

Franklin, D. W. & Wolpert, D. M. (2008). Specificity of reflex adaptation for task-relevant variability. *Journal of Neuroscience*, *28*(52), 14165–14175.

Franklin, D. W. & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*(3), 425–442.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*, *87*(1), 311–322.

Friedman, R. M., Hester, K. D., Green, B. G., & LaMotte, R. H. (2008). Magnitude estimation of softness. *Experimental Brain Research*, *191*(2), 133–142.

Frissen, I., Campos, J. L., Souman, J. L., & Ernst, M. O. (2011). Integration of vestibular and proprioceptive signals for spatial updating. *Experimental Brain Research*, *212*(2), 163–176.

Fujii, S., Kudo, K., Ohtsuki, T., & Oda, S. (2009). Tapping performance and underlying wrist muscle activity of non-drummers, drummers, and the world's fastest drummer. *Neuroscience Letters*, *459*(2), 69–73.

Gepshtein, S. & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*(6), 483–488.

Gescheider, G. A. (1974). Effects of signal probability on vibrotactile signal recognition. *Perceptual and Motor Skills*, *38*(1), 15–23.

Gilhodes, J., Roll, J., & Tardy-Gervet, M. (1986). Perceptual and motor effects of agonist-antagonist muscle vibration in man. *Experimental Brain Research*, *61*(2), 395–402.

Giordano, B. L. & Avanzini, F. (2014). Perception and synthesis of sound-generating materials. In M. Di Luca (Ed.), *Multisensory softness–Perceived compliance from multiple sources of information* (pp. 49–84). London, UK: Springer–Verlag.

Giordano, B. L. & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, *119*(2), 1171–1181.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). 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*, *36*(2), 462–476.

Giordano, B. L., Visell, Y., Yao, H.-Y., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *Journal of the Acoustical Society of America*, *131*(5), 4002–4012.

Goodwin, G. & Sin, K. (1985). *Adaptive filtering prediction and control*. Englewood Cliffs, NJ: Prentice–Hall.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*(2), 274–284.

Grassi, M., Pastore, M., & Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? *Acta Psychologica*, *143*(1), 96–104.

Grey, J. M. & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, *63*(5), 1493–1500.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, *146*(2), 161–171.

Guthrie, B. L., Porter, J. D., & Sparks, D. L. (1983). Corollary discharge provides accurate eye position information to the oculomotor system. *Science*, *221*(4616), 1193–1195.

Harris, C. M. & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784.

Hartmann, W. M. (1997). *Signals, sound, and sensation*. Woodbury, NY: Springer Science & Business Media.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 967–992.

Houben, M. M., Kohlrausch, A., & Hermes, D. J. (2004). Perception of the size and speed of rolling balls by sound. *Speech Communication*, *43*(4), 331–345.

Houben, M. M., Kohlrausch, A., & Hermes, D. J. (2005). The contribution of spectral and temporal information to the auditory perception of the size and speed of rolling balls. *Acta Acustica united with Acustica*, *91*(6), 1007–1015.

Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). Robpca: A new approach to robust principal component analysis. *Technometrics*, *47*(1), 64–79.

Hunt, K. & Crossley, F. (1975). Coefficient of restitution interpreted as damping in vibroimpact. *ASME Journal of Applied Mechanics*, *42*(2), 440–445.

ISO (2004). 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). Technical report, International Organization for Standardization, Geneva.

Itkowitz, B., Handley, J., & Zhu, W. (2005). The OpenHaptics™ Toolkit: A library for adding 3D Touch™ navigation and haptics to graphics applications. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems: World Haptics Conference*, (pp. 590–591)., Pisa, Italy.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*(8), 345–350.

Johansson, R. & Westling, G. (1988). Coordinated isometric muscle commands adequately and erroneously programmed for the weight during lifting task with precision grip. *Experimental Brain Research*, *71*(1), 59–71.

Jones, J. A. & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, *190*(3), 279–287.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*(6), 718–727.

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., & Yoshioka, T. (2003). Internal forward models in the cerebellum: fmri study on grip force and load force coupling. *Progress in Brain Research*, *142*, 171–188.

Keough, D. & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *Journal of the Acoustical Society of America*, *126*(2), 837–846.

Knill, D. C. & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*(5582), 846–848.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), e943.

Körding, K. P. & Wolpert, D. M. (2004a). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Körding, K. P. & Wolpert, D. M. (2004b). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(26), 9839–9842.

Körding, K. P. & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326.

Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2014). Audio–tactile integration and the influence of musical training. *PLoS One*, *9*(1), e85743.

Kunkler-Peck, A. J. & Turvey, M. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 279–294.

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*(27), 9351–9358.

LaMotte, R. H. (2000). Softness discrimination with a tool. *Journal of Neurophysiology*, *83*(4), 1777–1786.

Landau, L. & Lifshitz, E. M. (1981). *Theory of elasticity*. London, UK: Pergamon Press.

Landy, M. S., Banks, M. S., & Knill, D. C. (2011). Ideal–observer models of cue integration. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory cue integration* (pp. 5–29). NY: Oxford University Press.

Landy, M. S. & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*(9), 2307–2320.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Lederman, S. J. (1979). Auditory texture perception. *Perception*, *8*(1), 93–103.

Lederman, S. J. & Klatzky, R. L. (2004). Multisensory texture perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 107–122). Cambridge, MA: MIT Press.

Lederman, S. J., Klatzky, R. L., Morgan, T., & Hamilton, C. (2002). Integrating multimodal information about surface texture via a probe: Relative contributions of haptic and touch-produced sound sources. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, (pp. 97–104)., Orlando, FL.

Lemaitre, G. & Heller, L. M. (2012). Auditory perception of material is fragile while action is strikingly robust. *Journal of the Acoustical Society of America*, *131*(2), 1337–1348.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2006). *SAS® for mixed models* (2 ed.). Cary, NC: SAS Institute Inc.

Lutfi, R. A., Liu, C.-J., & Stoelinga, C. N. (2011). Auditory discrimination of force of impact. *Journal of the Acoustical Society of America*, *129*(4), 2104–2111.

MacLellan, M. J. & Patla, A. E. (2006). Adaptations of walking pattern on a compliant surface to regulate dynamic stability. *Experimental Brain Research*, *173*(3), 521–530.

Marhefka, D. W. & Orin, D. E. (1999). A compliant contact model with nonlinear damping for simulation of robotic systems. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans*, *29*(6), 566–572.

Marigold, D. S. & Patla, A. E. (2005). Adapting locomotion to different surface compliances: Neuromuscular responses and changes in movement dynamics. *Journal of Neurophysiology*, *94*(3), 1733–1750.

Martin, F. N. & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64–66.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, *115*(3), 1306–1320.

McAdams, S., Roussarie, V., Chaigne, A., & Giordano, B. L. (2010). The psychomechanics of simulated sound sources: Material properties of impacted thin plates. *Journal of the Acoustical Society of America*, *128*(3), 1401–1413.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192.

Miall, R. & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279.

Mitsuya, T., MacDonald, E. N., & Munhall, K. G. (2014). Temporal control and compensation for perturbed voicing feedback. *Journal of the Acoustical Society of America*, *135*(5), 2986–2994.

Moore, B. C. & Glasberg, B. R. (1983). Suggested formulae for calculating auditory–filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74*(3), 750–753.

Nowak, D. A., Hermsdörfer, J., Rost, K., Timmann, D., & Topka, H. (2004). Predictive and reactive finger force control during catching in cerebellar degeneration. *Cerebellum*, *3*(4), 227–235.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468.

Osborne, L. C., Lisberger, S. G., & Bialek, W. (2005). A sensory source for motor variation. *Nature*, *437*(7057), 412–416.

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, *22*(1), 46–49.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382–407.

Plack, C. J. & Moore, B. C. (1990). Temporal window shape as a function of frequency and level. *Journal of the Acoustical Society of America*, *87*(5), 2178–2187.

Pouget, A. & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, *3*, 1192–1198.

Rauschecker, J. P. & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800–11806.

Ronsse, R., Miall, R. C., & Swinnen, S. P. (2009). Multisensory integration in dynamical behaviors: Maximum likelihood estimation across bimanual skill learning. *Journal of Neuroscience*, *29*(26), 8419–8428.

Saijo, N., Murakami, I., Nishida, S., & Gomi, H. (2005). Large-field visual motion directly induces an involuntary rapid manual following response. *Journal of Neuroscience*, *25*(20), 4941–4951.

Saunders, J. A. & Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental Brain Research*, *152*(3), 341–352.

Saunders, J. A. & Knill, D. C. (2004). Visual feedback control of hand movements. *Journal of Neuroscience*, *24*(13), 3223–3234.

Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *Journal of Neuroscience*, *32*(14), 4913–4922.

Sedda, A., Monaco, S., Bottini, G., & Goodale, M. A. (2011). Integration of visual and auditory information for hand actions: Preliminary evidence for the contribution of natural sounds to grasping. *Experimental Brain Research*, *209*(3), 365–374.

Serrien, D. J. & Wiesendanger, M. (1999). Role of the cerebellum in tuning anticipatory and reactive grip force responses. *Journal of Cognitive Neuroscience*, *11*(6), 672–681.

Shadmehr, R. & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, *14*(5), 3208–3224.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, *33*, 89–108.

Shams, L. & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417.

Smith, M. A. & Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in huntington's disease but not cerebellar degeneration. *Journal of Neurophysiology*, *93*(5), 2809–2821.

Sober, S. J. & Sabes, P. N. (2003). Multisensory integration during motor planning. *Journal of Neuroscience*, *23*(18), 6982–6992.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004a). Congruency effects between auditory and tactile motion: Extending the phenomenon of cross-modal dynamic capture. *Cognitive, Affective, and Behavioral Neuroscience*, *4*(2), 208–217.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004b). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 330.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489.

Stocker, A. A. & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585.

Stoelinga, C., Hermes, D., Hirschberg, A., & Houtsma, A. (2003). Temporal aspects of rolling sounds: A smooth ball approaching the edge of a plate. *Acta Acustica united with Acustica*, *89*(5), 809–817.

Tan, H. Z., Durlach, N. I., Beauregard, G. L., & Srinivasan, M. A. (1995). Manual discrimination of compliance using active pinch grasp: The roles of force and work cues. *Perception & Psychophysics*, *57*(4), 495–510.

Toscano, J. C. & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869.

Tseng, Y.-w., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*, *98*(1), 54–62.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, *12*(10), 834–837.

Van der Linde, R. Q., Lammertse, P., Frederiksen, E., & Ruiter, B. (2002). The HapticMaster, a new high-performance haptic interface. In *Proceedings of Eurohaptics*, (pp. 1–5)., Edingburgh, UK.

Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York, NY: Springer–Verlag.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America*, *122*(4), 2306–2319.

Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *The British Journal of Animal Behaviour*, *2*(3), 89–94.

Warren, W. H. & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704–712.

Wei, K. & Körding, K. (2009). Relevance of error: What drives motor adaptation? *Journal of Neurophysiology*, *101*(2), 655–664.

Wei, K. & Körding, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Frontiers in Computational Neuroscience*, *4*(11).

Wei, K., Wert, D., & Körding, K. (2010). The nervous system uses nonspecific motor learning in response to random perturbations of varying nature. *Journal of Neurophysiology*, *104*(6), 3053–3063.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Welch, R. B. & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.

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

Wildes, R. & Richards, W. (1988). Recovering material properties from sound. In W. Richards (Ed.), *Natural Computation* (pp. 356–363). Cambridge, MA: MIT Press.

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. *Human Movement Science*, *26*(4), 511–524.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882.

Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7), 1317–1329.

Wolpert, D. M. & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, *22*(6), 996–1003.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.

Zarate, J. M. & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, *40*(4), 1871–1887.

Zwicker, E. & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2 ed.). New York, NY: Springer Berlin.