Perception of Action and Object Categories in Typical and Atypical Excitation-Resonator Interactions of Musical Instruments

Erica Y. Huynh[1,2†], Joël Bensoam[3], and Stephen McAdams[1,2]
[1] Music Technology Area, McGill University, Montréal, Québec, Canada
[2] Center for Interdisciplinary Research in Music Media and Technology, McGill University, Montréal, Québec, Canada
[3] Instrumental Acoustics Team, Institut de recherche et coordination acoustique/musique, Paris, France
[†] Corresponding author: erica.huynh@mail.mcgill.ca

Introduction

Our ability to recognize sound sources is automatic, yet very little is known about how this process works. Physical sources generate sounds that carry information about the object and material of the source and the action required to produce the sound. Listeners perceive impacted materials produced by the same action or material as more similar (Hjortkjær & McAdams, 2016) and can accurately identify the actions and materials of sound sources across broad categories (Lemaitre & Heller, 2012). A recent review suggests that listeners classify tones under the same category and rate them as more similar if they are played by similar excitation methods or instruments of the same family (Giordano & McAdams, 2010). These findings are often based on sound sets comprising tones from recorded orchestral instruments or synthesized versions of them, which are highly familiar in everyday listening. Siedenburg et al. (2016) collected dissimilarity ratings based on sounds that were familiar (i.e., recorded orchestral instrument tones) and unfamiliar (i.e., timbral transformations of recorded instrument tones that preserve acoustic properties). Perceived dissimilarity depended on an interplay of categorical (e.g., instrument family, excitation method) and acoustic (e.g., brightness) information. The goal of the current study is to directly examine how actions and objects are identified when they are combined in ways that are either typical or atypical of acoustic musical instruments. We used a synthesis paradigm that allowed for direct application of isolated actions to isolated objects. We assessed identification with two experimental methods: rating the resemblance of each stimulus to different actions (Experiment 1a) and objects (Experiment 1b), and explicit categorization of their actions and objects (Experiment 2).

Method

Stimuli. We used Modalys, a digital physical modelling platform, to synthesize stimuli that simulated three actions (bowing, blowing, and striking) and three objects (string, air column, and plate). Modalys allows for independent control of these actions and objects, so they can be freely associated, such that physically impossible sounds become possible with physically inspired modelling (Dudas, 2014). Thus, we combined each action with each object, creating nine classes of action-object interactions. Some interactions were typical of acoustic musical instruments (bowed string, blown air column, struck plate, and struck string), whereas others were atypical (bowed air column, bowed plate, blown string, blown plate, and struck air column). Interaction typicality was associated with the limitations of sound production in the physical world. For example, air columns and plates can be blown and struck, respectively; however, not many other actions set these objects into vibration in the physical world on a daily basis. Strings can be bowed and struck; but very rarely are they blown. Three exemplars for each action-object interaction were chosen through an exploratory approach to demonstrate the variability in their timbres (see Huynh, 2019).

Apparatus. Experiment 1 was conducted in the Perceptual Testing Lab at the Center for Interdisciplinary Research in Music Media and Technology (CIRMMT) at McGill University. It ran on a Mac Pro computer running OS 10.7 (Apple Computer, Inc, Cupertino, CA) and was displayed on an Apple Display 23-inch screen. We ran Experiment 2 in the Music Perception and Cognition Lab at McGill University in an IAC model 120act-3 double-walled audiometric booth (IAC Acoustics, Bronx, NY). The experiment ran on a Mac Pro computer running OSX (Apple Computer, Inc., Cupertino). In both Experiments 1 and 2, stimuli were presented over Seinnheiser HD280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany) and were amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA). The experiments were programmed in the PsiExp computer environment (Smith, 1995).

Procedure. Experiments 1a (N=41, 22 female, 19 male) and 1b (N=41, 24 female, 17 male) each comprised three blocks. Ratings in each block concerned only one type of action (1a) or object (1b). The order of the blocks was randomized for each participant. Each block had 27 trials—one for each stimulus (3 actions × 3 objects × 3 exemplars). In each trial, participants played a stimulus and rated its resemblance to the target action or object on a continuous scale from "not at all" to "completely". The order of stimulus presentation within each block was pseudo-randomized, such that two stimuli produced by the same action-object interaction were not presented in successive trials.

Experiment 2 (N=47, 35 female, 11 male, 1 other) involved a categorization task with two blocks. Each block concerned either action categorization or object categorization of the stimuli. The order of the blocks was randomized for each participant. Within each block, there were 27 trials. Depending on whether participants were categorizing the actions or objects of the stimuli, there were three boxes presented on the screen representing each category. Positions of the three boxes were randomized for each participant. For a given trial, participants played a stimulus and clicked the box corresponding to the action or object they thought produced it. The order of stimulus presentation within each block was pseudo-randomized in the same manner as in Experiment 1.

Results

Experiment 1. We averaged ratings for each action and object across the three exemplars of each action-object pair for each participant. We conducted two 3×3 repeated-measures Multivariate Analyses of Variance (MANOVAs), one for the action-resemblance ratings (Experiment 1a) and one for the object-resemblance ratings (Experiment 1b). For both MANOVAs, the independent variables were the action properties and object properties. We used Pillai's Trace, V, as the multivariate test statistic to accompany the F statistic, since it has been reported to be more robust to violations of assumptions (Olson, 1974).
For the action-resemblance ratings, we found significant within-groups effects of both the action properties (V=1.42, F(6,158)=65.11, p<.001, η_p^2=.71) and object properties (V=1.54, F(6,158)=87.20, p<.001, η_p^2=.77), as well as a significant interaction between actions and objects (V=0.99, F(12,480)=19.69, p<.001, η_p^2=.33). The different actions, objects, and their combinations influenced how listeners perceived the three actions. We conducted univariate analyses to test for the effects of the action-object interactions on each of the three action-resemblance ratings, since those interactions represent our stimuli. A conservative adjustment of the F statistic (i.e., epsilon) is reported to account for violation of the sphericity assumption. Significant interactions between actions and objects were revealed for: bowing ratings, F(3.68,147.12)=38.01, p<.001, ε=.92, η_p^2=.49; blowing ratings, F(3.77,150.76)=23.60, p<.001, ε=.94, η_p^2=.37; and striking ratings, F(2.58,103.20)=24.93, p<.001, ε=.65, η_p^2=.38.

For the object-resemblance ratings, the MANOVA revealed significant within-groups effects of: objects, V=1.56, F(6,158)=28.68, p<.001, η_p^2=.78; actions, V=1.04, F(6,158)=28.68, p<.001, η_p^2=.52; and their interaction, V=1.06, F(12,480)=21.82, p<.001, η_p^2=.35. Objects and actions that produced the sounds, as well as their interactions influenced listeners' perceptions of the objects. We further assessed this with univariate analyses to examine whether the different action-object interactions influenced resemblance ratings for each object. Accounting for the violation of the sphericity assumption, we report a conservative adjustment of the F statistic where appropriate. There was a significant interaction between action and object properties for: string ratings, F(3.32,132.88)=42.07, p<.001, ε=.83, η_p^2=.51; air column ratings, F(4,160)=34.72, p<.001, η_p^2=.47; and plate ratings, F(3.59,143.68)=33.62, p<.001, ε=.90, η_p^2=.46.

The significant interaction between action and objects for each of the action- and object-resemblance ratings are summarized in Figure 1. For stimuli representing typical action-object combinations, participants assigned the highest resemblance ratings to the actions and objects that actually produced the sounds. For example, bowed strings had the highest bowing and string ratings; similar patterns were also observed for blown air columns, struck strings, and struck plates. For the atypical interactions, listeners confused different actions or objects for one another. Bowed air columns and blown plates had the highest blowing and air column ratings. Additionally, bowed plates were perceived as struck plates and struck air columns were perceived as struck plates and struck strings. In these cases, listeners perceived either the correct action or object, seldom both. The correct property they perceived biased perception toward the complementary property that it most typically interacts with in acoustic musical instruments. An interesting case is the blown string, for which bowing, blowing, string, and air column ratings were highest. This may have to do with the fact that both bowed and blown sounds result from continuous excitations. As bowing is typically applied to strings and blowing is typically applied to air columns, participants confused these objects for one another.

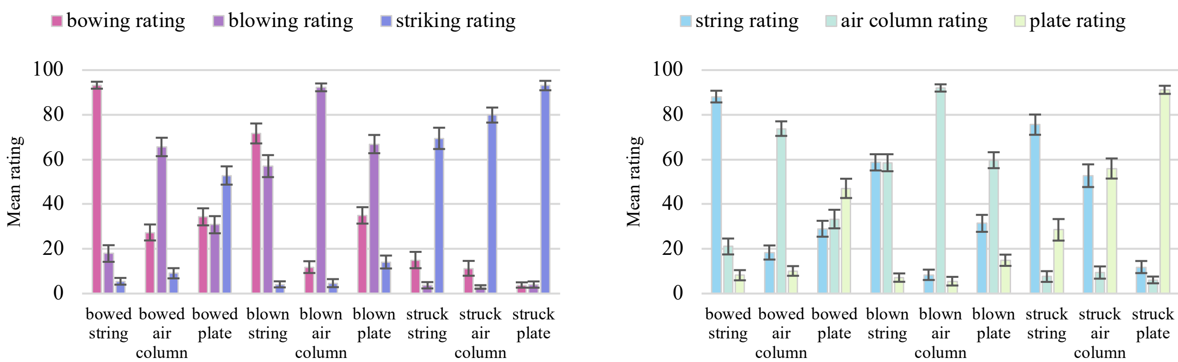

Figure 1: Mean action (left) and object (right) resemblance ratings for each action-object interaction. Error bars represent standard error of the mean.

Table 1: Categorization confusion matrix for the nine classes of action-object interactions. Stimuli are represented in the rows and response categories are represented in the columns. Typical stimuli are displayed in red font. Correct responses are indicated in bold.

	Action identification			Object identification		
Stimulus	**Bow**	**Blow**	**Strike**	**String**	**Air column**	**Plate**
bowed string	***131***	*6*	*4*	***134***	*5*	*2*
bowed air column	***17***	*123*	*1*	*9*	***123***	*9*
bowed plate	***59***	*23*	*59*	*16*	*26*	***99***
blown string	*91*	***48***	*2*	***85***	*55*	*1*
blown air column	*4*	***137***	*0*	*2*	***138***	*1*
blown plate	*41*	***98***	*2*	*19*	*93*	***29***
struck string	*24*	*0*	***117***	***129***	*0*	*12*
struck air column	*11*	*0*	***130***	*62*	***18***	*61*
struck plate	*4*	*0*	***137***	*3*	*4*	***134***

Experiment 2. Since this experiment involved explicit categorization of the actions and objects of the stimuli, we were able to further assess the ambiguities resulting from the resemblance ratings in Experiment 1. We computed confusion matrices for the categorization of actions and objects (Table 1). The values represent the number of times a stimulus was chosen as being produced by a certain action or object. Similar to Experiment 1, listeners correctly identified the actions and objects of stimuli produced by typical action-object interactions and confused different actions or objects for one another when they categorized atypical interactions. As in Experiment 1, bowed air columns and blown plates were categorized as blown air columns, and struck air columns were categorized as struck strings and struck plates. Listeners correctly categorized either the action or object and consequently perceived the complementary property it most typically interacts with. Listeners correctly categorized bowed plates as plates, but were confused between

bowing and striking. This may be because there was an impulsive sound arising from the bow's contact with the plate before the bowing occurs, which could be heard as a strike. For blown strings, which were rated as sounding like bowing, blowing, strings, and air columns in Experiment 1, listeners more decisively categorized them as bowed strings. Although they correctly identified the string, they were biased into perceiving bowing, the action most commonly applied to it.

Discussion

The current study demonstrates that action and object identification of nine types of action-object interactions depended on: (1) the familiarity with the interaction and (2) its perceived mechanical plausibility. Typical action-object interactions represented musical instrument families that listeners are familiar with: string instruments (bowed and struck strings), wind instruments (blown air columns), and percussive instruments (struck plates). Consequently, listeners are frequently exposed to the sounds these instrument families produce. Moreover, listeners were sensitive to the differences between impulsive (striking) and continuous (bowing, blowing) excitations, revealing that struck sounds were easily identified, but continuously excited sounds were often confused. For the atypical sounds, there was a general trend that listeners identified either the correct action or object and were consequently biased toward identifying the complementary property that most commonly interacts with it. This suggests that listeners seemed to interpret atypical interactions as conforming to a sound for which they already have mental models. Our findings indicate that it was either difficult for participants to perceive the actions or objects that interact atypically or that physically inspired modeling approaches cannot entirely convey what listeners simply do not have the mental models for. Although considering the timbre of a sound is essential for identifying its source properties, our findings suggest that our perceptions of sounds are not always accurate. Thus, novel sounds for which the actions or objects are difficult to identify are perceived as belonging to a category of sounds that has developed through a lifetime of exposure.

Acknowledgments

This research was financially supported through grants to Professor Stephen McAdams from the Natural Sciences and Engineering Research Council of Canada (NSERC). We would like to thank the Centre of Interdisciplinary Research in Music Media and Technology (CIRMMT) for providing us with the space to run participants.

References

Dudas, R. (2014). Modalys, Version 3.4.1 [computer software]. Paris: Institut de recherche et Coordination acoustique/musique.

Giordano, B. L., & McAdams, S. (2010). Sound source mechanics and musical timbre perception: Evidence from previous studies. *Music Perception*, *28*(2), 155–168.

Hjortkjær, J., & McAdams, S. (2016). Spectral and temporal cues for perception of material and action categories in impacted sound sources. *Journal of the Acoustical Society of America*, *140*(1), 409–420.

Huynh, E. (2019). *Bowed plates and blown strings: Odd combinations of excitation methods and resonance structures impact perception* (Unpublished master's thesis). McGill University, Montréal, Québec.

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.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, *69*(348), 894–908.

Siedenburg, K., Jones-Mollerup, K., & McAdams, S. (2016). Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in Psychology*, *6*, 1977.

Smith, B. K. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In: Wessel D. (ed), *Society for Music Perception and Cognition Conference*, (pp.83-84). Berkeley, University of California.