Predicting the timing of dynamic events through sound: Bouncing balls

Brian Gygi[a)]
Speech and Hearing Research, United States Department of Veterans Affairs Northern California Health Care System, Martinez, California 94553, USA

Bruno L. Giordano
Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland

Valeriy Shafiro
Department of Communication Sciences and Disorders, Rush University Medical Center, Chicago, Illinois 60612, USA

Anatoliy Kharkhurin
Department of International Studies, American University of Sharjah, Sharjah, United Arab Emirates

Peter Xinya Zhang
Department of Audio Arts and Acoustics, Columbia College, Chicago, Illinois 60605, USA

(Received 15 December 2014; revised 11 June 2015; accepted 15 June 2015; published online 22 July 2015)

Dynamic information in acoustical signals produced by bouncing objects is often used by listeners to predict the objects' future behavior (e.g., hitting a ball). This study examined factors that affect the accuracy of motor responses to sounds of real-world dynamic events. In experiment 1, listeners heard 2–5 bounces from a tennis ball, ping-pong, basketball, or wiffle ball, and would tap to indicate the time of the next bounce in a series. Across ball types and number of bounces, listeners were extremely accurate in predicting the correct bounce time (CT) with a mean prediction error of only 2.58% of the CT. Prediction based on a physical model of bouncing events indicated that listeners relied primarily on temporal cues when estimating the timing of the next bounce, and to a lesser extent on the loudness and spectral cues. In experiment 2, the timing of each bounce pattern was altered to correspond to the bounce timing pattern of another ball, producing stimuli with contradictory acoustic cues. Nevertheless, listeners remained highly accurate in their estimates of bounce timing. This suggests that listeners can adopt their estimates of bouncing-object timing based on acoustic cues that provide most veridical information about dynamic aspects of object behavior. © *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4923020]

[VMR] Pages: 457–466

I. INTRODUCTION

Successful navigation in everyday environments necessitates appropriate responses to diverse and dynamic sound sources (e.g., avoiding a collision with an approaching car, answering a door bell, turning to face a speaker behind you, or locating a leaking faucet). With visual information limited to objects and events in one's visual field, audition provides an important source of information about the environment through monitoring what is happening, where it is happening, and how it is happening in the space around the listener (Gaver, 1993). Naturally produced sounds can provide information about properties of the source objects, such as the material or internal structure, which may not be immediately obvious in the visual domain (e.g., Stoelinga, 2009, for rolling objects; for a review, see Giordano and Avanzini, 2014). In addition, for events in the auditory domain, humans need to be able to determine both the nature of the objects involved in an interaction (such as size, mass, hardness, shape) and the characteristics of the interaction between the objects (e.g., in Gaver's 1993 taxonomy of sounds, vibrating solids, aerodynamic sounds, and liquid sounds) in order to respond appropriately. However, little is known about human ability to estimate the timing of common real-world physical events based on the sounds they produce or factors that influence such timing estimates. To address this gap in current knowledge, the current study examined the accuracy of motor responses to a common and familiar dynamic event, bouncing pattern of a ball.

Previous investigations of human abilities to perceive the properties of everyday sound sources through sound alone focused largely on sound sources' static features. These have uncovered many auditory cues that specify aspects of the source, even across quite different exemplars of sounds from the same source in different listening situations. For example, listeners could judge the gender of a walker from the sound of his/her footsteps by focusing on spectral peaks and high-frequency components (Li *et al.*, 1991), utilize the changes in spectral level and spectral centroid over time to assess the hardness of a mallet used to strike a cooking pan (Freed, 1990), integrate frequency and

[a)]Electronic mail: bgygi@ebire.org

amplitude cues to determine the size of a ball and the thickness of the plate it was dropped on from the ensuing impact sound (Grassi *et al.*, 2013), and judge the fullness of a vessel from the change in fundamental frequency as water is poured in it (Cabe and Pittenger, 2000). Overall, these studies demonstrate that normal-hearing listeners are quite adept at tuning to the veridical information in acoustic signals in order to determine characteristics of sound-producing objects and events. This perceptual feat is especially impressive given that such information is generally scattered across a large variety of spectral and temporal acoustic cues.

Sound by its nature unfolds in time and, in contrast to the case for the static object properties described above, much less is known about human abilities to respond to temporal dynamics of real-world objects. In daily life, the ability to estimate and respond to dynamic behavior of real-world objects is essential to successful execution of many activities from crossing a busy street to swinging a tennis rocket to return serve, in which the timing of motor responses plays a crucial role. A great deal of research involving temporal aspects of sound perception of real-world sounds has focused on the beneficial or deleterious effects on speech comprehension of altering the rate of speech (e.g., Carver, 1973; Gygi and Shafiro, 2014) or fundamental studies of the temporal acuity and temporal resolution of the auditory system (see Moore, 2012, for an overview). Another area of research examined the mental chronology for artificially created temporal sequences that are often based on the abstracted temporal patterns in music. Large and colleagues, in their works, have emphasized the role of pulse, asserting that "…when humans organize complex temporal interactions, they synchronize—or more generally, entrain—pulse frequencies" (Large, 2008). They described pulse as a kind of endogenous periodicity, comprised of regularly recurring psychological events that arise in response to a musical rhythm. In this framework, pulse responds to onset timing of auditory events and can continue even in the absence of a stimulus (see, e.g., Large and Kolen, 1994; Large *et al.*, 2002).

To study timing accuracy for sounds, methods involving "sensorimotor synchronization" (SMS) are commonly employed, such as the synchronize-continue paradigm, used as far back as Stevens (1886). In this paradigm, participants listen to a periodic sequence, synchronize taps to the sequence, and continue tapping after the stimulus sequence is discontinued. This fairly simple method has yielded a great deal of knowledge about human abilities to follow sound events, termed entrainment (for a review, see Repp, 2005). Listeners can track temporally fluctuating sequences (Thaut *et al.*, 1998; Repp, 2001a), they respond quickly to phase perturbations of periodic sequences in which the phase is periodically reset (Large *et al.*, 2002; Repp, 2001b), and they tend to greatly underestimate the second of two consecutive time intervals under certain stimulus conditions (the "time-shrinking illusion"; e.g., Nakajima *et al.*, 2004). The findings from these studies have led to several complex models of timing (e.g., Jones and Boltz, 1989; McAuley and Kidd, 1995). However, the stimuli used to verify the models have largely been either acoustically controlled laboratory-generated sounds (e.g., pure tones or noise bursts) or musical tones. To date, the accuracy of listeners' motor responses to everyday sounds has not been systematically investigated despite their ubiquity.

Investigations of the time-varying characteristics of object interactions (i.e., sound-producing events) were largely limited to a handful of studies that did not involve specifically timed motor responses. These included discriminating between breaking and bouncing events, and judgments of the speed of rolling objects (Stoelinga and Chaigne, 2007). Another set of studies explored factors involved in dynamic judgments of sound source properties using continuously changing acoustic information, focusing on tau effects (tradeoffs between stimulus timing and spatial or pitch judgments, e.g., Shaw *et al.*, 1991; Sarrazin *et al.*, 2007), auditory looming (movement of a sound source toward a listener; for a review, see Neuhoff, 2004, Chap. 4), or estimating time when a vessel will be filled with water (Cabe and Pittenger, 2000). These studies demonstrate that humans are generally able to use dynamic acoustic information that produces accurate estimates of object behavior.

Unlike periodic sounds that have provided the basis for current entrainment models, most real-world sounds have varied timing (e.g., a walker speeding up or slowing down their gait) or have a damped periodicity. Thus, the entrainment models do not account well for events in which the timing might vary, which is the case for bouncing objects. Further, it is not clear how timing estimates vary for different types of objects for the same event, what acoustic cues may be most perceptually salient, and whether listeners can tune in to most informative acoustic cues to optimize response accuracy. These questions were investigated in the present study using the sounds produced by bouncing balls.

The temporally patterned sound of bouncing represents one of the more common dynamic environmental events. Bouncing can be described as a subset of impacts, in which the striking object is falling and is elastic enough not to shatter upon impact. The physics of bouncing objects have been quite well described (Cross, 1999), and information specifying the sound-producing event is available in the temporal as well as the spectral domain. In everyday activities, the temporal pattern of bouncing balls is commonly used in sports to inform responses such as picking up a ball or hitting it. In studying perception of cylinders of various materials involved in various actions, Lemaitre and Heller (2013) found that bouncing sounds allowed for the best identification of cylinder material and speculated that it was because a bounce consisted of repetitions of impacts, the timing of which was expected to reflect to the elastic properties of the cylinders' material.

However, the few studies that have directly examined the role of temporal information in perception of bouncing objects have been mixed as to its importance. Both Grassi *et al.* (2013) and Stoelinga (2009) found that temporal information was a rather poor predictor of performance in judging the size of bouncing balls, compared to spectral information. In contrast, Warren and Verbrugge (1984) examined the acoustic features that allowed listeners to discriminate bouncing from breaking objects, and found that temporal information was the best indicator of the type of sound-producing event. These studies had listeners form semantic judgments about properties of the sounding object after the completion of the event.

Little, however, is known about the ability to predict future events based on the temporally unfolding sound pattern, despite its importance in everyday life.

The present study examined the timing accuracy of motor responses based on the temporal dynamics of real-world sound sources and determined acoustic factors that inform those responses. Using the analysis-by-synthesis approach, we investigated (1) the relationship between mechanical properties of the sound producing objects (balls) and events (bouncing) and the resulting acoustic signals and (2) the acoustic signal cues used by the listener for estimating event timing. Manipulations included both the mechanical parameters of the sound-producing apparatus and the amount of information available to the listener to determine how these variables affect veridical estimates of temporal dynamics of the bouncing objects.

In experiment 1, the timing accuracy was examined for naturally produced bounces. Four different types of balls, representing different kinds of object materials, were dropped from three different heights, producing a wide range of temporal and spectral variation in the resulting bouncing sounds. Listeners heard a set number of bounces (from two to five, blocked by the number of bounces) and were asked to press a button when they expected the next bounce in the sequence to occur. Thus, listeners never heard the specific bounces whose timing they were asked to estimate, and needed to respond based solely on the timing of preceding bounces. Based on how well people respond to bouncing balls in the real world, it was predicted that listeners would be quite accurate in predicting the subsequent (unheard bounce) in a sequence. Physical models of bouncing events were then used to specify the types of acoustic cues listeners could use in making their predictions as to the timing of the next bounce.

Experiment 2 attempted to isolate which acoustic cues figured most strongly in listeners' judgments of the subsequent bounce. The technique of counterposing was used, in which stimuli were created that had the spectrum of one of the bouncing balls in experiment 1, but the bouncing pattern of another; these had conflicting spectral and temporal cues. Thus, the listeners' performance could indicate whether they were attending to the acoustic cues that were most predictive of the correct timing of bouncing events.

II. EXPERIMENT 1: TIMING ACCURACY FOR NATURAL BOUNCES

A. Methods

1. Stimuli

Stimuli were audio recordings of four different types of balls bouncing on a linoleum floor following a drop from one of three different heights: 100, 150, and 200 cm. The balls used were: (1) ping-pong ball, (2) basketball, (3) tennis ball, and (4) wiffle ball. The different types of balls were chosen to provide a sampling of materials whose physical properties would result in broad variation of spectral and temporal patterns produced during free bouncing. All recordings were made by a Zoom (Tokyo, Japan) H2 digital recorder held ~10 cm away from the location of the first bounce. Each recording was then truncated to create four different monaural audio files, each having a different number of bounces, i.e., the first two, three, four, or five consecutive bounces.

2. Design and procedure

Forty-eight unique stimuli were employed, which represented the combinations of three variables used: ball type (4), drop height (3), and number of bounces heard prior to motor response (4). The ball type and drop height stimuli were blocked by the number of preceding bounces (i.e., block 1 had two bounces, block 2 had three bounces, and so on) with each block containing 12 stimuli with the same number of bounces, but with various combinations of ball type and drop height. For each stimulus block, the 12 stimuli were presented 3 times, each time with a different randomization. The order for number of bounces was the same for all participants, starting with the two-bounce stimuli and ending with the five-bounce stimuli. This presentation order ensured that participants never heard the target bounce whose timing they were asked to estimate prior to making the responses. There were a total of 144 test trials in the test, preceded by 12 unscored practice trials that contained 2-bounce stimuli for each ball type and drop height.

Using E-Prime (Psychology Software Tools, Inc., Sharpsburg, PA) software, stimuli were presented diotically over KOSS (Milwaukee, WI) TD61 headphones at a comfortable listening level that participants could adjust during the practice trials. The presentation level was not changed after experimental trials began. Participants were informed that they would listen to sounds made by different balls as they bounced on a hard linoleum floor after being dropped from different heights, and were told what kinds of balls those were (but were not shown them). For all conditions, listeners were instructed on each trial to listen to a specified number of bounces and to press the response button at the time they expected the next bounce to occur (without actually hearing it). A Cedrus (San Pedro, CA) RB-610 serial response box was used for recording subjects' responses.

Prior to beginning the listening part of the test, participants answered a series of questions regarding their familiarity with each type of ball: (1) whether they played with each type of ball, (2) age they started playing, (3) for how many years, and, (4) how many days per week.

3. Participants

Participants were 27 undergraduate college students [average age 19.3 yr, standard deviation (SD) = 1.5; 16 females], enrolled in the Introduction to Psychology class at the University of Sharja. Participants received extra course credit for participation in the study. Data for two of the subjects were not used because of the large number of missing responses.

B. Results

Listener responses and stimulus characteristics were evaluated in three stages. First, listener response times (RTs) were compared to the actual physical times of the bounces. Second, acoustic parameters predicted by a physical model of bouncing events were compared to listener responses to determine their perceptually saliency. Finally, listener familiarity with the types of bouncing balls considered in this experiment was analyzed to determine its potential contributions to motor performance.

1. RT analysis

The RT data were preprocessed to discard both extremely early anticipations (cases in which the RT was less than the last heard bounce) and overly delayed responses (where the difference between the RT and the correct bounce time (CT) was more than two SD greater than the mean RT). This eliminated 196 responses across participants, or $\sim$5% of total responses.

Two main performance measures were computed individually for each subject: the mean deviation of RT from CT, or prediction error, and the SD of RTs across repetitions, which is an indicator of how reliable the subjects' estimates are. Note that both of these measures negatively scale with performance: a larger prediction error indicates greater *in*accuracy, and a larger RT SD value suggests greater *un*reliability. Furthermore, prediction error is signed, that is, it can be either positive (reflecting a mean RT that is greater than the CT, or response lag) or negative (in which the mean RT is less than the CT, or response lead).

On the other hand, the RT SD (which is unsigned, i.e., only takes positive values), reflects *response variance*, i.e., how much off the target listeners are on average in estimating the bounce timing. Thus, the relation of the two measures provides complementary information about the factors at play in listeners' responses. A small prediction error (responses very close to the actual CT value) combined with a large RT SD (which indicates a large variance in responses) suggests that listeners entrain well to the damped periodicity, but have a large amount of interference in the response process. A large mean RT deviation and a small RT SD means that the prediction error is large, but the noise in the response process is negligible.

Across ball types, bounce heights, and number of bounces heard, subjects were extremely accurate in their predictions as to when the next bounce would occur. There was a high correlation of RTs with CTs (minimum of the participant-specific Pearson RT-CT correlation = 0.966, $p(46) < 0.001$ for all participants). Surprisingly, rather than exhibiting an anticipation tendency, as could be expected based on previous rhythmic tapping research (Miyake, 1902; Repp, 2005), participants appeared to overestimate bounce time, i.e., their responses across all ball/bounce conditions significantly lagged behind the correct event by a mean of 58.13 ms, SD = 142.43 ms, one-sample $t(26) = 2.12$, $p = 0.044$. This prediction error equaled 2.58% of the CT, with a maximum of 196.05 ms in the five-bounce, 100-cm condition for the wiffle ball (which is a mean deviation of 7.13%). Otherwise, the RT accuracy overall is comparable with that found in other tapping studies; when listeners have to synchronize taps to a pulse, highly trained participants can achieve error rates as low as 2% (e.g., Pressing and Jolley-Rogers, 1997; Repp and Penel, 2002). It should also be noted that in the present experiment similar performance accuracy was reached with only a single tap and minimal practice. This suggests that listeners can synchronize their timing responses with dynamic temporal patterns of natural events even better than with series of pulses with arbitrary characteristics.

Next, three-way repeated measures analysis of variance (ANOVA) (ball type $\times$ drop height $\times$ N bounces) were performed on both the prediction error and response variance measures. There were significant main effects on prediction error by the number of bounces and ball type [$F(3) = 12.57$ and 8.00, respectively, $p < 0.001$], as shown in Fig. 1. Overall, the ping-pong ball had the most accurate responses and the tennis ball the least accurate. Contrary to what has been found in the majority of tapping studies (Repp, 2005), the prediction error for all ball types increased with the number of bounces heard (p-value for linear trend = 0.002). Possible reasons for and implications of these effects are included in Sec. II C. There was no main effect of height. However, there were significant two-way interactions of height with the number of bounces and ball type [$F(6) = 2.43$ and 3.09, respectively, $p \leq 0.029$]; these effects were complex and did not vary systematically with height, so they will not be discussed.

For response variance, there were three significant main effects: the number of bounces and ball type, $F(3) = 5.74$ and 2.74, respectively, $p < 0.05$, as well as drop height, $F(2) = 5.95$, $p < 0.005$. Among the balls, in contrast to prediction error, the ping-pong ball had the largest response variance; *post hoc* tests with Bonferroni corrections showed the mean SD across height and bounces was significantly greater for the ping-pong ball than either the tennis ball or the basketball. There was also a decrease in response variance with increasing height; the 200 cm height had significantly higher mean SD in *post hoc* tests than the 100 cm condition (recall there was no effect of height on prediction error). As with prediction error, the response variance became worse with a greater number of bounces (p-value for linear trend = 0.020), although the decrease was slight. *Post hoc* Bonferroni tests showed an unexpectedly low response variance for the tennis ball in the four-bounce condition.

To summarize, two experimental factors, ball type and number of bounces, had a significant effect on both performance measures, i.e., response prediction error and response variance, although the effect of specific balls differed between the two measures. In addition, drop height had a significant effect on response variance.

2. Physical modeling and analysis of behaviorally relevant acoustical information

Since the only information the listeners had regarding the physical dimensions of the stimuli was from the bouncing sounds, physical models of bouncing events were employed to determine which acoustic cues could be used to determine the timing of individual bounces. Then, listener responses were examined to determine which of the available acoustic cues were used by listeners to estimate bounce timing.

From a physical modeling standpoint, the information allowing prediction of the next bounce of any ball is

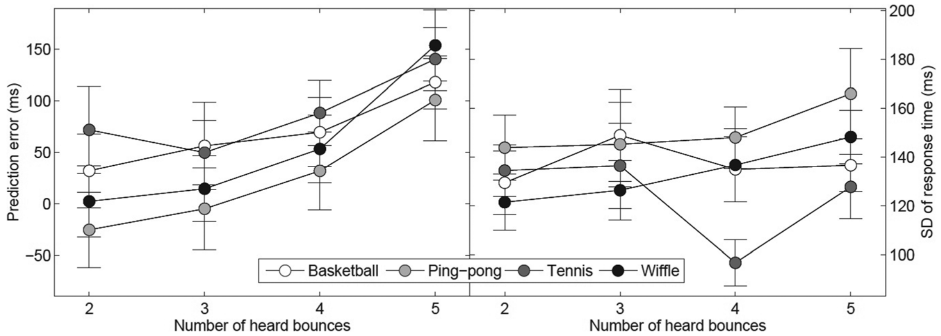

FIG. 1. Prediction error and response variance (SD of RTs) for each ball type across bounce conditions.

available in the acoustics. Assuming that on each bounce the percentage of potential energy that is dissipated is a constant, the ratios of time intervals between bounces are constant, as are the ratios of the heights of the bounces (and thus the amplitudes of the impact). Therefore, if listeners have heard three bounces (i.e., two bounce intervals labeled t_2 and t_3), then the fourth bounce can be predicted solely on the basis of the timing of the bounces, which we will term the timing model

$$t_4 = \frac{t_3^2}{t_2}.$$

However, if the listener heard only two bounces, to determine the temporal location of the third bounce, the model needs to include the amplitude levels of the bounces, which decrease in constant steps. Including this yields a formula for predicting the third bounce

$$t_3 - 10^{\Delta L/20} t_2,$$

where L is the level of the bounce. This will be referred to as the level + timing model. In this case, it is not obvious whether L should refer to the peak instantaneous amplitude of the bounce segment or to the root-mean-square (rms) in a window containing the bounce. Thus, different acoustic cues provide veridical information about the bounces for the two vs more than two bounce patterns. A number of various-sized time windows were used and a 10-ms window was regarded as most accurately capturing the bounce segments. Applying these formulas using the onset times and amplitudes of the bounces (both instantaneous peak and 10 ms window from the start of the segment) yielded extremely close correspondences for the third, fourth, fifth, and last bounces for each ball type. Overall, the prediction using the bounce timing alone was more accurate, with a mean deviation from the actual onset of the sound of 8.944 ms over all ball types and bounces. However, the prediction error of the timing information varied across bounces—for the fourth bounce (the first one for which relative time could provide a prediction), the mean deviation was 14.124 ms, but for the sixth bounce it was 4.507 ms, indicating an improvement with the number of bounces. This is in contrast to human performance, which declined with the number of bounces.

For the combination of interbounce interval and level, the overall mean deviation when using the peak intensity was 73.13 ms and when using the 10 ms averaging window was 53.076 ms. Again, the correspondence of the physical model to the actual onset was worse for the early bounces. On the third bounce, for which the timing model cannot offer a prediction, the peak intensity model had a mean deviation of 89.295 ms and the 10 ms averaging window yielded a mean deviation of 70.393 ms. Thus, if we compare the listeners to ideal observers, they should be much less accurate on the third bounce, when they only can use timing through level; however, as noted earlier, the humans were best in estimating the timing of the third bounce. In addition, it should also be noted that on the third bounce, when only the level + timing model can offer a prediction, the human listeners outperform the model regardless of whether using peak amplitude or rms in a 10 ms window.

Following comparison of human performance with predictions of analytical event models, we examined the relation between the available acoustical information and behavioral responses. Temporal, energetic, and spectral properties of the sounds were calculated from the time-varying specific loudness (Glasberg and Moore, 2002). For each bounce, the following variables were extracted: (1) temporal location of loudness peak, (2) peak loudness, and (3) spectral center of gravity (SCG) at loudness peak. Based on the parameters of the two models, two groups of these acoustical features were considered to be relevant: (1) the temporal location, loudness, and SCG of the last heard bounce and (2) for each of these three features, the difference in value between the last and second-last heard bounce. The extent to which each of these acoustical features accurately specified CT of the physical bounce was assessed using an information-accuracy score, defined as the absolute Spearman correlation between CT and the target acoustical feature (Giordano *et al*., 2010); the score for each feature is listed in Table I.

A multiple rank-regression model incorporating all six acoustical features was used to predict the RTs for each of the participants; in each case, the RTs were accurately predicted by these acoustical features (minimum variance in the RT ranks accounted for by the acoustical features was 88%). The best predictor by far was the temporal location of the last heard bounce, information accuracy = 0.99 (which is very close to the performance of an ideal observer). This yielded a much greater information-accuracy score than the difference between the last two-heard bounces, whereas the other two measures, SCG and loudness, are comparable for the last heard and the last two-heard bounces. Also listed in Table I are the mean Spearman rank correlation (ρ) and

TABLE I. Across-participants mean correlation, ρ, and partial correlation, ρ_p, between the behavioral estimates of bounce time and each of the acoustical features. Also reported are the information-accuracy scores (Inf. Acc.) measured with the stimuli investigated in experiment 1. An information-accuracy score of zero corresponds to the chance-level performance of an ideal listener who produces random estimates, whereas an information-accuracy score significantly higher than zero is interpreted as indicating better-than-chance rating performance.

	Experiment 1		Experiment 2—unmodified		Experiment 2—hybrids		
	ρ	ρ_p	ρ	ρ_p	ρ	ρ_p	Inf. Acc.
Time_L	0.98**	0.98**	0.80**	0.80**	0.76**	0.76**	0.99**
Loudness_L	−0.44**	0.15*	−0.38**	−0.15**	−0.34**	−0.17**	0.46**
SCG_L	0.19**	0.09*	0.21**	0.35**	−0.10**	−0.17**	0.22
Time_Δ	−0.56**	0.34**	−0.28**	−0.31**	−0.36**	−0.36**	0.56**
Loudness_Δ	0.50**	−0.04	0.06**	−0.28**	0.04*	−0.04	0.51**
SCG_Δ	−0.14**	0.07*	−0.04*	0.06*	−0.11**	0.15**	0.13

Note: * = p-value < 0.05, ** = p-value < 0.001, L = last-heard bounce, Δ = difference between last and second-last heard bounce.

mean Spearman rank partial (ρ_p) correlation between each of the acoustical features and the RTs. Although all the features correlated significantly with the RTs, the time of the last-heard bounce had by far the strongest correlations, mean ρ and $\rho_p = 0.98$. Next, behavioral weight was computed from the absolute correlation between each acoustical feature and listeners' bounce time estimates. These are plotted in Fig. 2 along with the informational accuracy for each feature (the corresponding values for experiment 2, described below, are plotted alongside). It thus appears that listeners based their perceptual judgments on acoustic features that are most informative in terms of their information-accuracy scores, thus, approximating the behavior of the ideal observer.

3. Effect of familiarity on performance

Subjects' responses to the questions regarding familiarity with each type of ball were examined and coded categorically to enable ANOVA of prediction error and response variance for bounce timing responses. The age at which users first played with the ball was coded as 0—never played with the ball, 1—started at age 7 or less, 2—started at above age 7. The numbers of years played with the ball was coded as 0—never played with the ball, 1—played with the ball for seven years or less, 2—played with the ball for more than seven years. Finally, the number of days of the week playing with the ball was coded as 0—do not play at all during week, 1—play one day or more during the week. Familiarity remained at yes/no.

The prediction error and response variance were subject to main-effects ANOVA with the familiarity codes as categorical factors. For prediction error, the only significant main effect was for the number of years played ($p < 0.05$). There was no significant difference between 0 (no experience with the ball) and 1 (played with the ball for seven years or less), both of which had a mean prediction error of 65–72 ms. However, for subjects who had played with the ball for more than seven years, the mean prediction error was −70 ms. Thus, the prediction error was of the same magnitude, but anticipatory rather than lagging. Apparently, experience with the balls did not enable subjects to more accurately predict the bounces, but it led to a different type of response strategy.

However, the familiarity scores had a much stronger influence on response variance. There was a main effect for all the familiarity scores, and the effects were all in the direction of lower SD of RTs. Thus, it seems that greater familiarity with the ball makes the responses more consistent, possibly by enabling the development of a more coherent neural model of the physics of the ball-bouncing event. This possibility will be explored further in Sec. IV.

C. Discussion

The highly accurate timing estimates of unheard bounces suggest that listeners have an accurate and reliable mental model of the dynamic behavior of bouncing events. The listeners' model appears quite supple and is able to accommodate a number of object and event parameters such as ball type, drop height, and the number of preceding bounces. Although responses to familiarity ratings suggest some contribution of experience with bouncing balls to the consistency of responses, accuracy of timing estimates was strikingly high even when listeners did not have experience with these specific types of balls. Given the richness of the stimuli, the model appears to involve a wide range of acoustical properties (temporal, energetic, and spectral), and to

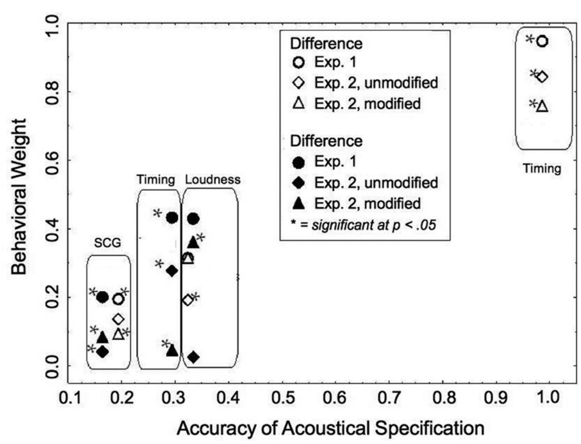

FIG. 2. Behavioral weight and informational accuracy for each stimulus group in both experiments. The open symbols refer to the values derived from the last-heard bounce, and the filled symbols are from the difference between the last-heard bounce and the second-to-last heard bounce (see Sec. II B 2 for explanation).

consider both temporally local information (from the last-heard bounce) and information about the temporal evolution of the acoustical structure (from the last- and second-last bounces). While multiple regression analyses suggest highest contributions to performance accuracy from the temporal location of the last-heard bounce, performance worsened with a greater number of bounces. This is in contrast to the findings from the vast majority of tapping studies, in which performance tends to improve with the number of external events the listener is presented with. However, in tapping studies, listeners make continuous tapping responses to discrete and typically periodic sounds. This gives them sufficient time to synchronize their responses with the auditory stimulus and self-correct when necessary (Repp, 2005). In contrast, the present procedure required only a single tap, while listeners had to keep track of the number of bounces presented to know when to make a response. As the number of bounces increased, this may have increased the memory load, diverting processing resources from tracking the bounce timing. Additional factors that may explain the relationship between performance accuracy and the number of preceding bounces will be considered in Sec. IV.

Despite the behavioral relevance of a number of acoustical features, listener responses to dynamic bouncing events appear to be more strongly influenced by temporal features that most accurately predict bounce time. To further test whether listeners can selectively tune in to acoustic information that is most predictive of CT, in experiment 2 the acoustic cues were mixed so that they contained contradictory information. The timing of the bounces in each stimulus file in experiment 2 was altered to correspond to the bounce timing pattern of another ball for each of the three drop-height conditions. In these "hybrid" bounces, contrasting the temporal information contained in the bouncing pattern of one ball with the spectral and energetic information of another ball could indicate which acoustic cues carried a greater perceptual weight in listener estimates.

III. EXPERIMENT 2: TIMING ACCURACY FOR HYBRID BOUNCES

A. Methods

1. Stimuli

Stimuli in experiment 2 were based on the two- and three-bounce audio files used in experiment 1 for each height and type of ball. However, the timing of the bounces in each stimulus file in experiment 2 was altered to correspond to the bounce timing pattern of another ball for each of the three drop-height conditions by digitally editing the sound files. Specifically, the bounce timing pattern of the ping-pong ball was imposed on bouncing events of the tennis ball, wiffle ball, and basketball, while the bounce timing pattern of the wiffle ball was imposed on the bouncing events of the ping-pong ball. In addition to the time-altered hybrid stimuli, two- and three-bounce stimuli from experiment 1 were used as experimental controls. To clearly delineate the stimulus manipulations, the hybrid conditions will be referred to as ping-pong/wiffle, wiffle/ping-pong, basketball/ping-pong, tennis/ping-pong, with the first ball name denoting the sounding object in each bounce and the second ball name denoting the bounce timing pattern.

The hybrid stimuli were checked against the original stimuli to ensure accuracy of spectral and temporal information. The loudness and SCG of the temporally modified sounds were very close to what was measured in the original stimuli, average across-bounces difference was 0.29 dB and 0.09%, respectively, SD = 0.32 dB and 0.9%, respectively. As a second step, we assessed whether the temporal location of the hybrid-sound bounces corresponded to temporal location of the target original-sound bounces. The temporal patterning of the hybrid-sound bounces corresponded closely to that of the target original-sound bounces, mean absolute across-bounces difference = 8.5 ms, or 0.7%, SD = 9.2 ms.

2. Design and procedure

As in experiment 1, stimuli were blocked by the number of bounces with each block containing 24 unique stimulus files corresponding to different types of ball (4), different heights (3), and different bounce timing patterns (2: the original vs hybrid). In the course of experiment 2, each block was presented three times for each of the two number-of-bounce conditions (first two-bounce than three-bounce blocks) for a total of 144 test stimulus trials. Prior to the test trials proper, participants responded to a practice two-bounce blocks of 24 trials and answered the questions about their previous exposure to each ball. In all other respects, the procedure of experiment 2 was identical to that of experiment 1.

3. Participants

Participants were a new group of 30 college students (average age 19.7 yr, SD = 1.9; 24 females) with no previous experience with the experimental stimuli. They were similarly enrolled in an Introduction to Psychology course and received course credit for participation.

B. Results

As in experiment 1, the RT data were preprocessed to discard both anticipations and overly delayed responses; 205 responses, or 5.2%, were thus discarded. Then, the prediction error (mean deviation) and response variance (mean SD of RTs) measures were calculated. Once again, across ball types, bounce heights, number of bounces heard, and stimulus type conditions (original vs hybrid), subjects were quite accurate in their predictions as to when the next bounce would occur, as shown in Fig. 3. The correlation between listeners' RTs and the CT was again quite high, $r = 0.997$. Also, as in experiment 1 listeners showed a tendency to overestimate the bounce time, both for the hybrid (mean deviation = 59.9 ms; SD = 339.4 ms) and original (mean dev. = 33.7 ms; SD = 421.8 ms) stimuli. A paired t-test showed the mean prediction error for the original stimuli to be significantly lower than for the hybrid, $t(26) = 3.76$, $p < 0.001$. The mean prediction error as a percentage of the

CT was fairly low for both stimulus types, 3.3% for the hybrid stimuli and 1.7% for the original.

The original and hybrid data in all stimulus conditions were entered into a four-way repeated-measures ANOVA, with stimulus type, height, ball type, and number of bounces as factors. For prediction error, aside from the main effect of stimulus type, discussed above, there were main effects of ball type and number of preceding bounces. Two-way interactions were found for ball type and number of bounces, stimulus type and ball type, and height and ball type. Significant three-way interactions were also noted for stimulus type, height, and ball type, as well as stimulus type, ball type, and number of bounces.

Overall, the mean prediction error for both the original and hybrid stimuli in experiment 2 are higher than for the two- and three-bounce stimuli in experiment 1 (mean $= 20.46$ ms, SD $= 546.79$). In addition, the largest prediction error in experiment 2, in the three-bounce 200 cm ping-pong/wiffle condition, 539 ms (27.2% of the CT), was much greater than that found in experiment 1 (196 ms, 6.69% of CT).

However, the diminished accuracy performance in experiment 2 was largely confined to the two-bounce condition, as demonstrated in Fig. 4, which plots prediction error by the number of bounces for the original and hybrid conditions, along with the comparable data from experiment 1. In the two-bounce condition, planned comparisons showed that the prediction error for the original stimuli in experiment 1 was significantly less than both the hybrid and original stimuli from experiment 2 ($F = 8.015$, $p = 0.007$). Thus, the confusion introduced by mixing the hybrid stimuli with the original stimuli seemed to affect performance on the original stimuli in the two-bounce condition. Given contradictory acoustic information, listeners might have been less certain about their responses to the two-bounce stimuli which contained limited temporal cues. On the other hand, in the three-bounce condition performance on both the original and hybrid stimuli were not significantly different from the original stimuli from experiment 1. It seems that with longer exposure to the bounce pattern which provided additional timing cues, the subjects learned that the energetic cues were not reliable and shifted their focus to acoustic parameters that were more effective in predicting CTs.

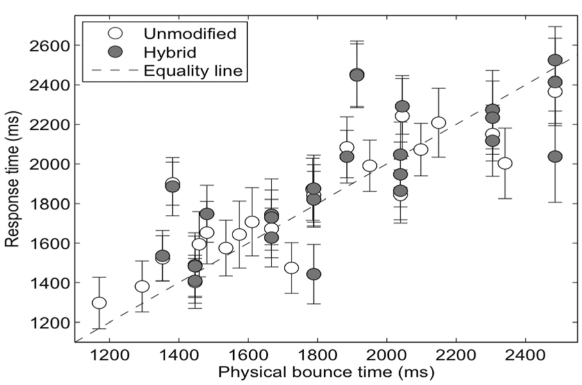

FIG. 3. Plot of the physical bounce time against subjects' RTs for experiment 2.

In the case of ball type, in experiment 2 performance on the hybrid stimuli was comparable to that of the original stimuli, but in an unusual way, demonstrated in Fig. 5, which plots the original and hybrid results by ball type and drop height. Although there was a significant three-way interaction between these factors, the original and hybrid plots resemble each other in a striking way. The basketball/ping-pong ball hybrid matched almost exactly the ping-pong ball results alone in that both had quite a large jump in mean deviation for the 200 cm height. Similarly, the ping-pong/wiffle ball hybrid tracked very closely the wiffle ball alone data; however, in this instance, there was a large *negative* dip in mean deviation for the 200 cm height. Planned comparisons between the hybrids and the stimulus providing the bounce pattern failed to yield a significant difference for either pair. In both these cases, the ball that provided the temporal bouncing pattern seemed to dominate the perception. However, the wiffle/ping-pong ball and tennis ball/ping-pong ball hybrids both yielded quite low mean deviations, with no difference between the drop heights, indicating greater effect of spectral profile. Possible reasons for this are examined in Sec. IV.

The response variance data for experiment 2, i.e., the mean SD of the prediction error, exhibit a different pattern; unlike the comparable data for experiment 1, the only significant effect was an interaction of stimulus type, ball and height, $F(6,48) = 139.35$, $p < 0.001$. This appears to be largely due to an unusually low SD for the basketball/ping-pong ball hybrid at the 200 cm height. There was no effect of the number of bounces, although in experiment 2 there were only the two- and three-bounce conditions, which did not show a significant difference in experiment 1 either. A mixed-effect ANOVA with the response variance data from experiment 1 revealed no overall main effect of experiment and planned comparisons showed no significant pairwise difference between the control data from experiment 1 and either the control or hybrid data from experiment 2.

C. Discussion

The results of experiment 2 show that listeners can adapt different strategies in estimating the timing of future events

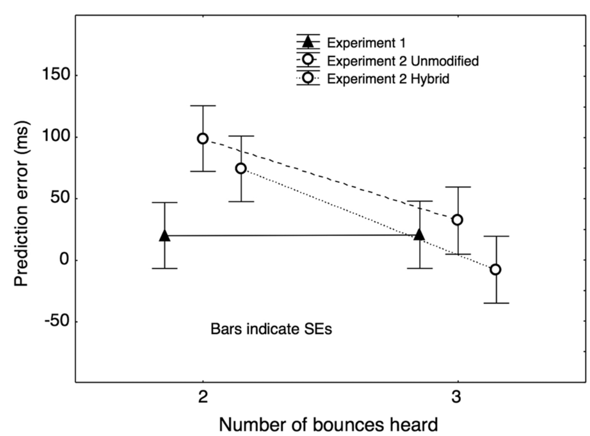

FIG. 4. Prediction error in experiment 1 and the two stimulus conditions in experiment 2 by number of bounces heard.

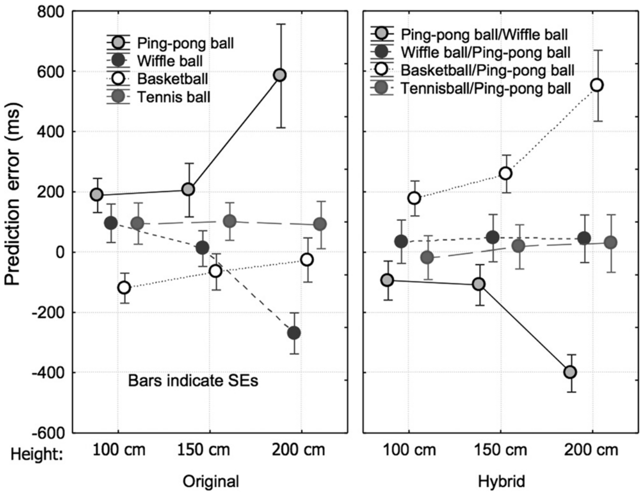

FIG. 5. Prediction error by ball type and drop height for the control and hybrid stimuli in experiments 1 and 2.

and base their responses on the acoustic cues that provide the most veridical acoustic information about bouncing-object behavior. In doing so, listeners appear to use information present within and across trials. In the first block, the two-bounce condition, for which only half of the sounds followed a pattern that could be expected in a free fall from a particular height, listeners performed less well on both the hybrid stimuli and the original stimuli than did listeners in the comparable condition in experiment 1. With only two bounces, the physical modeling suggests that listeners could not rely on temporal pattern cues alone to estimate the timing of the third bounce, and had to also attend to the loudness and spectrum of the bouncing objects. It appears the hybrid condition introduced a great deal of uncertainty about which acoustic cues were most informative for estimating the following bounce, which degraded performance for the original stimuli more than the hybrid stimuli.

In contrast, in the three-bounce condition, the performance with both the hybrid and control stimuli improved significantly, so much so that the prediction error for the hybrid stimuli in that block was comparable to the prediction error for the original stimuli from experiment 1. This suggests that when some acoustic parameters are not informative (such as loudness or SCG, which were from a different ball type for the hybrid stimuli), listeners learned to focus on other more informative properties, such as the timing of previous bounces.

IV. GENERAL DISCUSSION

Sound by its nature unfolds in time and, in contrast to the case for the static object properties described above, little is known about human abilities to estimate temporal dynamics of real-world objects. In daily life, the ability to predict and respond to dynamic behavior of real-world objects is essential to successful execution of many activities from crossing a busy street to swinging a tennis racket. In these, the timing of motor responses plays a crucial role. The findings presented here suggest that listeners have a well-developed ability for extracting complex attributes of familiar everyday physical events from the acoustics alone. In this study, listeners proved highly accurate at predicting the timing of a subsequent bounce of a variety of different balls and different drop heights simply by hearing the previous bounces. Physical modeling suggested that when three or more bounces are heard, highly accurate performance in this task could be achieved by focusing on the change in bounce timing; specifically, the difference T_3 (the time of the last bounce) - T_2 (second-to-last bounce) compared to $T_2 - T_1$ (the third-to-last bounce). When only two bounces are heard, correct predictions could be made using a temporal feature, T_2, and a more spectrally based one, the amplitude level difference between T_2 and T_1.

As a test of the importance of these acoustical features to listeners, hybrid stimuli were created in which the temporal and spectral properties were counterposed: they had the temporal patterning of one ball and the bounce sounds of a different ball. In the two-bounce condition, listeners were much less accurate for the hybrid stimuli than for the control (i.e., original) stimuli. However, when hearing three or more bounces, listeners performed as well in the hybrid condition as in the original condition. Thus, it seems that listeners are able to adjust their listening strategies to adapt to different situations.

Listeners' ability to accurately predict the timing of the following bounces, which can be conceptualized as a mental model, has likely developed from experience with the balls. That users reporting greater experience with a particular ball showed a decreased response variability strongly implies that with experience they formed more stable and precise mental models of the balls' bouncing behavior. As noted, these models seem to rely largely on temporal information, which is why performance is comparable for the original and hybrid stimuli in the three-bounces-heard condition. This is not a process of simple template matching; the mental model must be more flexible than this to allow for different stimulus conditions. Since bouncing produces a damped periodic signal with intervals becoming shorter throughout, it does not lend itself well to application of more purely rhythmic models, such as entrainment (Large and Kolen, 1994) or synchronization (McAuley and Jones, 2003), which rely on even intervals and phase locking, or the time-shrinking illusion (Nakajima *et al.*, 2004) in which prior shorter time intervals cause a later longer interval to be perceived as shorter than it really is.

The correlation analyses also suggest that listeners relied only secondarily on the SCG; however, the poorer performance in the two-bounce condition for the hybrid stimuli in experiment 2 indicates that the SCG does play a significant role in listeners' judgments, at least when they only hear two bounces. It is possible that when there are few bounces, listeners used the SCG to identify the type of ball being dropped, which will determine how bounce timing information is used. However, when there are more bounces listeners can either ignore the SCG information, or the additional information from the subsequent bounce(s) makes the SCG information superfluous.

In experiment 1, performance worsened in the four- and five-bounce conditions: this result is opposite to what would be expected if participants were learning the specific bouncing patterns they were hearing, and to what has been found, as noted above, in the majority of tapping studies. It also is contrary to predictions of the physical model, which were actually less accurate for the two- and three-bounce conditions. Although the findings presented here do not point to any clear reason for this effect, one possibility is that the performance decrease is due to attentional and memory limitations that arise when listeners have to hold in short-term memory four or five bounces. Another possibility is that the finding is simply an artifact of procedural response proclivities. Since the later bounces come more rapidly, the intervals to be judged are shorter and factors, such as internal noise in the response process, are going to have a potentially greater effect on RTs.

In conclusion, the results of this work are generally consistent with the ecological approach to perception. Listeners focus on the aspects of the stimulus that enable the most accurate response to complete a task; in this case, using the timing of the bounces to predict the temporal location of a subsequent bounce. Listeners are also good at ignoring irrelevant or misleading information, such as the spectral information in the case of hybrid stimuli. The ability to use veridical acoustic information to determine behaviorally relevant object and event properties likely comes from exposure to common sound sources in the course of everyday life.

ACKNOWLEDGMENTS

We would like to thank Dr. Sejal Kuvadia for her help with making the recordings for this project.

Cabe, P. A., and Pittenger, J. B. (**2000**). "Human sensitivity to acoustic information from vessel filling," J. Exp. Psychol. Hum. Percept. Perform. **26**, 313–324.

Carver, R. P. (**1973**). "Effects of increasing the rate of speech presentation upon comprehension," J. Educ. Psychol. **65**, 118–126.

Cross, R. C. (**1999**). "The bounce of a ball," Am. J. Phys. **67**, 222–227.

Freed, D. (**1990**). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," J. Acoust. Soc. Am. **87**, 311–322.

Gaver, W. W. (**1993**). "What in the world do we hear? An ecological approach to auditory event perception," Ecol. Psychol. **5**, 1–29.

Giordano, B. L., and Avanzini, F. (**2014**). "Perception and synthesis of sound-generating materials," in *Multisensory Softness—Perceived Compliance from Multiple Sources of Information* (Springer, London, UK), pp. 49–84.

Giordano, B. L., Rocchesso, D., and McAdams, S. (**2010**). "Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability," J. Exp. Psychol. Hum. Percept. Perform. **36**, 462–479.

Glasberg, B. R., and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds," J. Audio. Eng. Soc. **50**, 331–342.

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

Gygi, B., and Shafiro, V. (**2014**). "Spatial and temporal modifications of multitalker speech can improve speech perception in older adults," Hear. Res. **310**, 76–86.

Jones, M. R., and Boltz, M. (**1989**). "Dynamic attending and responses to time," Psychol. Rev. **96**, 459–491.

Large, E. W. (**2008**). "Resonating to musical rhythm: Theory and experiment," in *The Psychology of Time* (Emerald Publishing Group, West Yorkshire, UK), pp. 189–231.

Large, E. W., Fink, P., and Kelso, J. A. S. (**2002**). "Tracking simple and complex sequences," Psychol. Res. **66**, 3–17.

Large, E. W., and Kolen, J. F. (**1994**). "Resonance and the perception of musical meter," Connect. Sci. **6**, 177–208.

Lemaitre, G., and Heller, L. M. (**2013**). "Evidence for a basic level in a taxonomy of everyday action sounds," Exp. Brain Res. **226**, 253–264.

Li, X. F., Logan, R. J., and Pastore, R. E. (**1991**). "Perception of acoustic source characteristics: Walking sounds," J. Acoust. Soc. Am. **90**, 3036–3049.

McAuley, J. D., and Jones, M. R. (**2003**). "Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing," J. Exp. Psychol. Hum. Percept. Perform. **29**, 1102–1125.

McAuley, J. D., and Kidd, G. R. (**1995**). "Temporally directed attending in the discrimination of tempo: Further evidence for an entrainment model," J. Acoust. Soc. Am. **97**, 3278.

Miyake, I. (**1902**). "Researches on rhythmic activity," Studies From the Yale Psychological Laboratory **10**, 1–48.

Moore, B. C. J. (**2012**). *An Introduction to the Psychology of Hearing*, 6th ed. (Brill Academic Publishers, Leiden, NE), pp. 1–442.

Nakajima, Y., ten Hoopen, G., Sasaki, T., Yamamoto, K., Kadota, M., Simons, M., and Suetomi, D. (**2004**). "Time-shrinking: The process of unilateral temporal assimilation," Perception **33**, 1061–1079.

Neuhoff, J. G. (**2004**). *Ecological Psychoacoustics* (Elsevier Academic, London), pp. 1–350.

Pressing, J., and Jolley-Rogers, G. (**1997**). "Spectral properties of human cognition and skill," Biol. Cybern. **76**(5), 339–347.

Repp, B. H. (**2001a**). "Processes underlying adaptation to tempo changes in sensorimotor synchronization," Hum. Movement Sci. **20**, 277–312.

Repp, B. H. (**2001b**). "Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization," J. Exp. Psychol. Hum. Percept. Perform. **27**, 600–621.

Repp, B. H. (**2005**). "Sensorimotor synchronization: A review of the tapping literature," Psychon. Bull. Rev. **12**, 969–992.

Repp, B. H., and Penel, A. (**2002**). "Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences," J. Exp. Psychol. Hum. Percept. Perform. **28**, 1085–1099.

Sarrazin, J.-C., Giraudo, M.-D., and Pittenger, J. B. (**2007**). "Tau and kappa effects in physical space: The case of audition," Psychol. Res. **71**, 201–218.

Shaw, B. K., McGowan, R. S., and Turvey, M. T. (**1991**). "An acoustic variable specifying time to contact," Ecol. Psychol. **3**, 253–261.

Stevens, L. (**1886**). "On the time sense," Mind **11**, 393–404.

Stoelinga, C. (**2009**). *A Psychomechanical Study of Rolling Sounds* (VDM Verlag Dr. Müller, Saarbrücken, GE), pp. 1–208.

Stoelinga, C., and Chaigne, A. (**2007**). "Time-domain modeling and simulation of rolling objects," Acta. Acust. Acust. **93**, 290–304.

Thaut, M. H., Miller, R. A., and Schauer, L. M. (**1998**). "Multiple synchronization strategies in rhythmic sensorimotor tasks: Phase vs. period correction," Biol. Cybern. **79**, 241–250.

Warren, W. H., and Verbrugge, R. R. (**1984**). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," J. Exp. Psychol. Hum. Percept. Perform. **10**, 704–712.