The Effects of Temporal Action-Sound Congruence on Evaluations of Conductor Quality

Cory D. Meals1*, Steven J. Morrison2 and Deborah A. Confredo3

Abstract
Vision serves a fundamental role in the human experience of musical performance. In conducting, this particular heuristic influences both expressive and coordinative aspects of musical activity. Ensemble conductors present a special case of musical gesture, as their activities are coordinative rather than directly sound-producing. While the influence of vision on evaluations of musical expressivity has been well studied, less attention has been paid to the temporal aspect of conductors’ gestures. Given anecdotal observations of a flexibly congruent relationship between conductor gesture and ensemble response and the ability of entrainment to promote preference, we theorize that alterations to natural action-sound congruence in conductor-to-ensemble settings may influence evaluations of conductor quality. Naturalistic performance video of five conductors was left intact or adjusted to an audio- or video-lead condition by a percentage of each excerpt tempo (intact, ± 15%, ± 30%) and fully crossed into stimuli orders. Participants were asked to rate the quality of the conductor, the ensemble, and the performance overall using a Likert-type scale bound by “poor” and “excellent.” Our results indicate that any offset, whether audio- or video-led, resulted in a lower level of conductor quality than intact, unaltered performance. While our effect size was small ($\eta^2 = .02$), participant ratings reinforce the role of action-sound congruence on observers’ perceptions and overall evaluation of conductors’ activities.

Keywords
Action-sound congruence, conducting, ensemble performance, entrainment, music

Introduction
Vision serves a fundamental role in the human experience of music performance. While its propensity for broad audio/visual influence is well established (McGurk & MacDonald, 1976), the feature-rich context that music presents highlights the pervasive universality of its influence and the wide variety of manifestations it takes. Examples of this influence are as varied as expressivity (Dahl & Friberg, 2007; Davidson, 1993), tone duration (Schutz & Lipscomb, 2007), and performance quality of individuals (Tsay, 2013) and groups (Tsay, 2014). Even within these, vision’s influence extends to both the mechanical (Fendrich & Corballis, 2001) and affective or expressive (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2006) within musicians’ performance behaviors. The observed perceptual interaction of sound and sight is situated within a narrowed view of embodied music cognition (Leman & Maes, 2014), where action’s effect on music is central to the perception and understanding of it. Interestingly, the overwhelming majority of music performance bolsters this view by providing a directly linked action-sound pairing of performer gesture to instrument response save for one case: that of the conductor.

Seen as central to the function of instrumental ensembles, the conductor is unique in considerations of music as their sole offering is non-auditory. Through gesture alone,
they convey both the affect and temporal coordination of ensemble musicians, leading them by their deployment of motions that are thought of as alternately “navigational” or “evocative” (Leinsdorf, 1981), with the latter steering affect and the former guiding precision. While some offer that conductors are, often as not, “selected from among the worst pianists to be found, or [those that] cannot play the pianoforte at all” (Berlioz, 1926, p. 257), this individual’s contributions are universal within Western instrumental ensemble performance, lending credence to the conductor’s role in the function of the ensemble. To wit, former conductor of the New York Philharmonic Alan Gilbert offers that “There is a connection between the gesture, the physical presence, the aura that a conductor can project, and what the musicians produce” (2012). This link between conductor and performer is further bolstered by the physical location of the conductor—often directly in the view of all performers at the front of the ensemble—which allows performers to easily see their contributions.

Despite their soundless contribution, the conductor’s gestures are intended to inform a wide range of musical function and sonic outcomes within the ensemble (Rudolf, 1980). Pedagogically focused conducting texts are explicit in centralizing the conductor’s coordinative role and treating expressivity as a necessary extension (see Pasquale, 2008 for a review). Often, this is presented as “mere time-beating is not enough; the appropriate gesture for each musical expression must be mastered before we can speak of conducting” (Rudolf, 1980, p. xiv). While the communication of time can be viewed pejoratively, its necessity is embedded in the requisite nature of its hierarchical position articulated by Rudolf. Interestingly, this implicitly graduated bifurcation of time (e.g., navigation) into affect (e.g., evocation) in conductor gesture is also supported by empirical reports. The temporal coordination of the ensemble performers (D’Ausilio et al., 2012; Hilt et al., 2013; Luck & Ne, 2008; Luck & Sloboda, 2007; Luck & Toivianien, 2006) as well as the communication of affect within the work performed (Kumar & Morrison, 2016; Morrison, Price, Geiger, & Cornacchio, 2009; Morrison, Price, Smedley, & Meals, 2014) have both been shown significant influence by the conductor’s gesture.

Expressive gesture influences perception and subsequent evaluation of music performance with near universality (Dahl & Friberg, 2007; Platz & Kopiez, 2012). Noted at the level of the individual musician (Davidson, 1993; Tsay, 2013; Vines et al., 2006), it has also been identified in the gestures of instrumental (Montemayor & Silvey, 2019; Morrison et al., 2009, 2014; Silvey, Wacker, & Felder, 2017) and choral (Morrison & Selvey, 2014; Napoles, 2013) conductors. The ability of these soundless gestures to filter and influence perceptions even in the absence of supporting sonic information bolsters the statement that “the visual component is not a marginal phenomenon in music perception, but an important factor in the communication of meaning” (Platz & Kopiez, 2012, p. 75) as it applies to this context. Within and around this, however, the tendrils of ensemble coordination are tightly woven in the activity of the conductor.

Ensemble coordination is effected by the use of a prescribed gestural vocabulary that prominently features points of elision between intersecting horizontal and vertical vectors (Luck & Toivianien, 2006). Within conducting literature, each point of elision is referred to as an ictus (Green & Gibson, 2004; Rudolf, 1980), serving as the indicator of isochronous pulse within the conductor’s gesture (Clayton, 1986). Performers extract temporal information through visual beat induction, where temporal signals are extracted from a combination of gestural acceleration and deceleration paired with changes in the value of parameters that define the gesture’s trajectory (Luck & Sloboda, 2008). Put another way, performers extract time-bearing information from both the changes in direction, as well as the rate of that change, within the conductor’s patterns. Once encoded, this serves as a point of reference for each performer’s analytical and future-oriented attending (Jones & Boltz, 1989), allowing for the coordination of individual-level performance of complex musical material with others via the conductor.

This process joins a range of coordinative interperformer signaling present in small groups (Chang, Livingston, Bosnyak, & Trainor, 2017; Glowinski et al., 2019; Schiavio & Hoffding, 2015) and in larger ensemble configurations (Goebl & Palmer, 2009; Hilt et al., 2013; Timmers, Endo, Bradbury, & Wing, 2014) that complement the conductor’s communication. D’Ausilio and colleagues (2012) investigated leadership in orchestral performance through analysis of nonverbal communication along both conductor-to-musician and musician-to-musician channels. Using audio- and motion capture data, they investigated sonic and kinematic interactions both between performers and with the conductor, identifying a temporally infused “driving force” of each in a given performance. From this they noted a flexibility between individual musicians and the conductor, finding an emergent hierarchy in higher-rated performances where the driving force of the conductor was more salient. In this, the authors noted that “aesthetic appreciation of music orchestras’ performance was associated to the concurrent increase of conductor-to-musician influence and a reduction of musician-to-musician information flow” (p. 4). The ensemble, then, presents a dynamic communication network within which the conductor’s gesture is highly salient to, but by no means exclusively responsible for, musical coordination.

This musical ecosystem represents an instance of social, multi-focal entrainment. Rhythmic detection, production, as well as phase and period correction (Clayton, 2012; Levitin, Grahn, & London, 2018; Philips-Silver, Aktilis, & Bryant, 2010; Repp, 2005) are present throughout the ensemble’s activities, highlighting the potential for coordinative flexibility rather than strict synchrony. Taken alongside the findings of D’Ausilio and colleagues (2012) and
bolstered by related work by Glowinski and colleagues (2019), the externalities of temporal detection, production, and correction actually represent two overlapping systems of entrainment: one between musicians and the other between the conductor and each musician.

These complementary and coexistent systems represent an example of social (Philips-Silver et al., 2010) entrainment existing in an asymmetrical, or one-to-many (Clayton, Sager, & Will, 2005), relationship that is crossmodal in its function (Philips-Silver & Trainor, 2007). The balance of that complexity varies between ensembles, performances, and sometimes within given works (D’Ausilio et al., 2012; Glowinski et al., 2019; Luck & Toivianen, 2006). Musicians’ ancillary, or non-sound-bearing, gestures vary both in quantity and complexity depending on the musical context in which they are deployed, as highlighted by Glowinski and colleagues (2019). Similarly, conductors have been anecdotally observed to alter their gestural vocabulary depending on the size of the group they are conducting, often adhering to an inverse relationship between complexity and size of ensemble (Acklin, 2009; Leinsdorf, 1981; Rudolf, 1980). Interestingly, the resultant flexibility within the entrained system appears to take on unique qualities across some performance contexts, as the congruence of the conductor to ensemble action-sound relationship is at times seen as, in fact, incongruent (Bell, 2004).

This phenomenon has been noted by many performers, audience members, and conductors (Bell, 2004; Bennett, 2017; Johnson, 2014), but has received limited scholarly or empirical attention. The presence of variability in synchronization with conductor gesture (Luck & Sloboda, 2008) suggests that correction of phase and period are present in this relationship (Repp, 2005). While anecdotal accounts speak to flexibility contrafactually, as conductor-to-ensemble relationships are not identified as qualitatively orthogonal, examples of severe misalignment seem to produce a negative or incredulous response (Arhombus, 2014). Given the role and importance of the conductor’s rhythmic signaling and entrainment’s ability to engender emotion (Trost, Labbé, & Grandjean, 2017), this presents an interesting case for consideration. Mirrored activities as simple as tapping fingers (Hove & Risen, 2009) or the coordinated movement of an appendage (Sparenberg, Topolinski, Springer, & Prinz, 2012) produce preference between individuals. Evidence also suggests that organized dance and movement (Dean, Byron, & Bailes, 2009; McNeill, 1995) report generally similar outcomes, even in group settings. When viewed alongside the development and persistence of crossmodal integration of body movement and auditory encoding (Philips-Silver & Trainor, 2007; Spence, 2011), and the propensity toward extracting coordinative information from gesture (Luck & Sloboda, 2007), the potential for temporal perturbations to influence evaluations of conductor efficacy becomes salient.

Anecdote suggests that the temporal incongruity in conductor gesture to ensemble response is common in orchestral settings (Roberts, 2017), but that it is a variable property that depends on the musical content being performed and the preferences of the conductor themself (Johnson, 2014). Given the flexibility inherent in entrained systems, especially those as large as a musical ensemble, and the propensity for entrainment to promote affinity (Sparenberg et al., 2012) and preference (Hove & Risen, 2009), temporal action-sound congruence outside a range of affordance may impact evaluations of conductor quality, regardless of their skill in conducting. To this end, we selected a large ensemble context where synchrony is almost exclusively the expectation in conductor-to-ensemble coordination (e.g., wind band), and posed the following research question: “Does the temporal displacement of conducting gesture from its sonic correlates have any bearing on evaluations of conductor efficacy and quality in a large ensemble setting?”

**Method**

**Participants**

Participants (N = 112) were recruited from the graduate and undergraduate populations of two music schools located in the Eastern and Midwestern U.S. All participants had one or more years of ensemble performance experience (M = 6.10 years, SD = 6.28 years); data collection was conducted through an online platform (Psychdata.net) and was carried out using university-based computer networks in accordance with Human Subjects protocols.1 Participants who did not complete the survey instrument (n = 2) were omitted from analysis. Remaining participants (n = 110) reported more instrumental/band performance experience (M = 8.35 years, SD = 6.04 years) than choral (M = 6.48 years, SD = 7.00 years) and orchestral (M = 3.46 years, SD = 4.54 years). Additionally, participants reported fewer years of conducting experience (M = 4.54 years, SD = 1.27 years) than musical performance (M = 6.10 years) experience across all ensemble types.

**Stimuli**

Stimuli were taken from intact performance videos collected from five experienced conductors (2 female). In all videos, conductors appeared from the perspective of the performers, with a variance of approximately 15 degrees to the right or left of ensemble center. The similarity of evaluations of conductor gesture from the performer’s perspective to the audience’s perspective has been demonstrated in similar studies (Price & Mann, 2011). Across all videos, performing musicians and audience members are viewable in the fore- and background, respectively. Video quality across collected samples met or exceeded baseline requirements for high definition (720p HD, 720...
x 480 pixels). Video artifacts were addressed using color correction and cropping where they did not detract from the focus of attention (e.g., automatic white balancing and cropping to center conductor in viewable field). Audio resolution for all videos was determined to be 16-bit WAV (CD quality) or better and represented professionally recorded audio captured during the performance in question.

Two musically cohesive excerpts per conductor (fast/slow) were chosen from the collected videos (fast: $M_{\text{fast}} = 153$ beats per minute (BPM), $SD_{\text{fast}} = 15.4$ BPM; slow: $M_{\text{slow}} = 76$ BPM, $SD_{\text{slow}} = 27.18$ BPM). These 10 samples (see Table 1) were selected for their consistent tempo throughout each excerpt and for the balanced presence of consistent and clear gestural communication within a fully intact and coherent musical phrase, although tempo variance among slow excerpts was greater than that of fast. Consideration was also given to the presence of contextually appropriate, expressive gesture shown by conductors. Excerpt length ($M = 29.2$ seconds) was consistent with previous research in conductor evaluation (Morrison et al., 2009, 2014).

Intact stimuli videos were manipulated within Adobe Premiere Pro (Adobe, Inc.). Excerpts were extracted and manipulated to allow for experimental manipulation through the separation and adjustment of audio and video tracks. This created a set of excerpt iterations encompassing an intact sample with no manipulation, a $\pm 30\%$ audio/visual offset condition, and a $\pm 15\%$ audio/visual offset condition for a total of 50 manipulated stimuli. The degree of offset reflected consideration of previous research highlighting similar offset procedures in conducting task synchronicity (Luck & Toiviainen, 2006), synchrony in music performance (Vatakis & Spence, 2006), and error detection in musical performance (Danz & Janyan, 2009). Offsets in the current study were chosen to allow for potential differences to emerge within given tempo conditions (fast/slow) rather than in a fixed offset relationship (as used by Vatakis & Spence, 2006 and discussed by Repp, 2005) where excerpt tempo might be a less salient variable. Much of the existing literature where static offset values were employed focuses on temporal order judgments (see Repp, 2005), a related but not identical task to the focus of our investigation.

These manipulations did not alter content quality or speed, but rather shifted the audio to either a lead (-) or lag (+) condition as compared to the intact video (see Figure 1). This temporal distance, measured in milliseconds (ms), was determined by analyzing the tempo of each excerpt and determining the quantity of offset required ($M_{\text{fast}} = \pm 138$ ms ($\pm 30\%$), $\pm 69$ ms ($\pm 15\%$); $M_{\text{slow}} = \pm 262$ ms ($\pm 30\%$), $\pm 131$ ms ($\pm 15\%$)). For clarity, we labeled stimuli by using the position of the audio stream related to the video where “minus” indicated audio positioned before video and “plus” indicated audio positioned after video.

The generated offsets were theorized to fall within a range of manipulation that would remain above the threshold of just noticeable difference (JND) by capitalizing on the ensemble experience of our participants ($M = 6.1$ years) and their familiarity with the multiple affordances present within ensemble performance (e.g., co-performer entrainment, conducting practice). While the offset selection exceeds the findings of others (Drake & Botte, 1993; Large, 2008), and may be seen to extend beyond what Weber’s law suggests regarding a perceivable change in stimulus intensity, it is important to remember that in many settings participants may experience a priming effect due to the nature of the task undertaken. In this experiment, musically experienced participants were asked to evaluate conductor quality rather than detect synchrony—and the flexible nature of conductor-to-ensemble timing paired with the complexities of ensemble performance constitute a perceptual affordance that is otherwise untested in this specific setting. Additionally, the complex web of communication in naturalistic performance combined with the crossmodal nature of the excerpts supports a perceptual adaptation of temporal incongruity similar to that reported by Keetels and Vroomen (2008), as well as Large’s observation that “there is a tendency for pulses to gravitate toward event onsets in a way that produces (approximate) synchrony when a stimulus rhythm is purely periodic” (2008, p. 193).

Conversely, stimulus manipulations fall outside the critical range of crossmodal integration reported by Fendrich and Corballis (2001). In their study, temporal fusing of combinations of visual flashes and auditory clicks and flutters presented in close temporal proximity to one another was observed below a threshold of 50 ms. Mean offset values across present stimuli were designed to fall outside this range. We theorize that this provides participants with variance that is meaningful enough to elicit a significant effect in evaluations yet not so severe that manipulations are perceived as a product technical error.

A 45-second “Please Respond” prompt was added to the end of each excerpt and each was rendered with high-definition audio and video (H.264, Apple, Inc.). From these stimuli five fully crossed orders were organized where no
conductor, offset value, or excerpt condition (fast/slow) was seen consecutively. Stimuli were individually uploaded to YouTube (Alphabet, Inc.) as unlisted videos.²

**Results**

A significant main effect was found for ensemble evaluation such that ratings for the slow tempo condition \((M = .124, SE = .042)\) were significantly lower than the fast tempo condition \((M = .308, SE = .036)\) \((Z = 11.25, p < .001, \eta_p^2 = .04)\). The main effect for offset condition was not significant \((F(4,1090) = 2.35, p > .05)\). However, the interaction effect between tempo condition and offset condition was significant \((F(4,1090) = 6.89, p < .001, \eta_p^2 = .02)\), and post hoc comparisons using the Tukey HSD test indicated that the mean score for slow excerpts in the +30\% offset condition \((M = -.190, SE = .103)\) was significantly different than those in the -15\% slow \((M = .288, SE = .085)\), intact slow \((M = .273, SE = .107)\), -30\% fast \((M = .305, SE = .085)\), intact fast \((M = .314, SE = .078)\), +15\% fast \((M = .471, SE = .075)\), and the +30\% fast \((M = .421, SE = .079)\) offset conditions. The persuasiveness of this interaction is understandable, however, since ensemble repertoire was not controlled for in the study design and both fast \((M = 152\ BPM, SD = 16.21\ BPM)\) and slow \((M = 76\ BPM, SD = 27.1\ BPM)\) tempo conditions varied by conductor and piece.

There was a significant main effect for conductor evaluation such that all offset conditions (audio- and video-lead) were rated lower than intact stimuli \((F(4,1090) = 4.668, p < .001, \eta_p^2 = .02)\). Post hoc comparisons using the Tukey HSD test revealed a significant difference between the intact \((M = -.087, SE = .075)\) and the +30\% \((M = -.467, SE = .078)\) conditions. The main effect for tempo condition was not significant \((F(1,1090) = .004, p > .05)\), and there was no significant interaction effect between offset condition and tempo condition in conductor evaluations.

As can be seen in both overall mean participant ratings (see Figure 3a) and participants ratings by excerpt condition (Figure 3b), an inverted U shape exists broadly across evaluations of both conductor and ensemble, where ratings were lower in any offset condition and conductor ratings were lower than ensemble ratings. Differences in both conductor and ensemble evaluations were observed even when considering the idiosyncratic styles and gestural magnitude demonstrated by each individual conductor.

---

² Meals et al., 5

---

**Data Analysis**

Collected participant ratings were standardized by participant to account for individual variance in scale usage using mean rating and standard deviation. Ratings were then organized by offset condition (+ 30\%, ±15\%, intact), as this was the central independent variable. Data exploration revealed that the “overall” rating generally reflected the mean of conductor and ensemble evaluations, which led to its exclusion from analysis. The remaining rating categories of “Conductor” and “Ensemble” were analyzed for differences by excerpt condition (fast/slow), by offset condition, and interactions between differences by excerpt and offset condition were investigated using a 2x5 factorial analysis of variance.
Discussion

Conducting represents a unique element of music performance, as its gesture does not have a direct sonic consequent. Instead, gesture serves as a coordinating signal from conductor to ensemble that conveys both temporal and affective information. Within this, musicians entrain to a number of rhythmic signals in ensemble performance (D’Ausilio et al., 2012), with the conductor’s central location allowing for their communication to remain salient to all performers. From that, this signaling combines with other layers of cooperative information exchange (e.g., complementary and imitative) to leaven the predictive qualities of musical joint action (Hilt et al., 2013). The conductor provides temporal information to the ensemble and these time-bearing gestures share a flexible relationship to the ensemble’s activity and output (Bell, 2004; Johnson, 2014), although there exists a clear directionality to the embedded information from conductor to ensemble (Hilt et al., 2013). As entrainment has been shown to engender preference (Hove & Risen, 2009) and affinity (Sparenberg et al., 2012), we sought to determine the aesthetic effects of manipulation to temporal action-sound congruence in conductor gestures.

Our results indicate that differences exist within both conductor and ensemble evaluations where gestural action-sound congruence has been manipulated, although the qualities of these differences are themselves different. Interestingly, a broad view of our data revealed that the
majority of offsets were rated lower regardless of orientation of action to sound (e.g., video- or audio-lead) (see Figure 3a). Exceptions to this are revealed through division of ratings by tempo condition (e.g., fast and slow), but the generally inverted U shape to the distribution—especially with regard to slow excerpt evaluations—is intriguing. In this same vein, no participants reported their videos appearing to be altered, suggesting that the manipulation of the stimuli fell within an acceptable range of affordance for the participants, who are experienced ensemble musicians ($M_{XP} = 6.10$ years). While this range of affordance varies from other literature on participants’ identification of “just noticeable difference” (JND) (Drake & Botte, 1993), the degree to which ensemble performance integrates multiple, multimodal streams of communication opens the possibility for a wider range of affordance in this setting (Hilt et al., 2013).

In conductor evaluations, all offset conditions were rated significantly different than intact video, with the majority rated lower than intact. The same did not hold true for ensemble evaluations, although a similar, inverted U shape appears in slow excerpt ensemble evaluations that generally mirrors their analogous conductor evaluations (Figure 3b). That evaluations of slow excerpts were generally rated lower than fast stands to reason, as offset values were greater due to the experimental protocols (as suggested by Large, 2008). A complicating aspect to this, however, is that the accompanying audio also features material of a longer duration across the musical spectrum (e.g., note duration, interonset interval, harmonic motion), which may serve as a mediating factor in these evaluations. Additionally, a consistent trend for conductor ratings to be lower than those of the ensemble pervaded evaluations, which is notable in its own right and bears further investigation, potentially building upon the framework offered by Kumar and Morrison (2016).

These findings raise two critical questions. First, would the response profile between slow and fast excerpts be more similar if each was adjusted using the same absolute (rather than relative) magnitudes of deviation? If so, this would reveal either (1) a potential duration-general window of “affordance” within which observers reconcile disjunct audio and video music performance information (as offered by Drake & Botte, 1993), or (2) a “blind spot” that lies

Figure 3. a-b. Standardized participant ratings (a) overall and by (b) tempo condition (fast/slow).
outside the range of discernment of decontextualized stimulus but may be facilitated by a more content-rich perceptual environment, an eventuality supported by work in crossmodal perceptual binding (Engel, Roelfsema, Fries, Brecht, & Singer, 1997).

Second, given a direct comparison between intact and adjusted excerpts, would a more apparent positive response to intact performances emerge for fast excerpts? Likewise, would a correspondingly more pronounced positive response emerge for slow excerpts? It is difficult to determine this from the data at hand, as stimuli were not organized in a fashion to centralize direct comparison, instead focusing on the influence of offset quantity. In contrast to fixed offset values, such a finding would confirm that the adjustments used in the present study were sufficiently discernible and that the responses observed reflect an active accommodating of temporal misalignment.

Ensemble performance presents us with a complex, multimodal communication network in which conductors are a critical, influential, but not isolated node. The results of this study suggest that the temporal congruence of their gesture bears consequence not only in their coordinative role, but in evaluations of quality in their performance, as well as the ensemble’s. Additionally, these results reinforce existing findings within synchrony preference and attendant prososocial behavior and reward center activity. Although this study adds broad support to the crossmodal influence of sight and sound in musical contexts, further research is needed to foster a more robust understanding of the interactions that dominate the dense mesh of activity, communication, and interaction within ensemble music.

Author contribution
CM and SM researched literature and conceived the study. DC was involved in study design and joined CM in participant recruitment and data collection. All authors reviewed and edited the manuscript and approved the final version.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this paper.

ORCID ID
Cory D. Meals  https://orcid.org/0000-0003-1224-5104

Notes
1. Psychdata.net collects IP address information from all participants, but anonymity was achieved via blanket IP address assignment within university networks utilized.
2. The nature of “unlisted” videos on media platforms such as YouTube restricts access to only those with a direct link to the media. This differs significantly from private videos that, while essentially invisible to web searches, are often password protected, un-embeddable, and therefore inappropriate for use through online survey forms such as those available through Psychdata.net.

References
Acklin, A. I. (2009). The effect of conducting on ensemble performance: A “best evidence” synthesis (Doctoral dissertation). Retrieved from https://diginole.lib.fsu.edu/islandora/object/fsu:168192/datastream/PDF/view;The
Arhombus. (2014, January 12). Why do elite orchestras play behind the conductor’s beat? [Online forum]. Retrieved from https://www.reddit.com/r/classicalmusic/comments/1v1is6/why_do_elite_orchestras_play_behind_the/cenv8yq?utm_source=share&utm_medium=web2x
Bell, L. (2004). Conducting: When orchestras are used to playing behind the beat [Online forum thread]. Retrieved from http://archive.choralnet.org/view/221793
Bennett, J. (2017, September 17). Why do orchestras seem to play behind the beat? WQXR Blog. Retrieved from https://www.wqxr.org/story/why-do-orchestras-play-behind-beat/
Berlioz, H. (1926). The orchestral conductor: Theory of his art. New York, NY: Carl Fischer, Inc.
Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. Proceedings of the National Academy of Sciences, 114, E4134–E4141. doi:10.1073/pnas.1617657114
Clayton, M. (1986). Coordination between players in musical performance (Doctoral dissertation). Edinburgh, UK: University of Edinburgh.
Clayton, M. (2012). What is entrainment? Definition and application in musical research. Empirical Musicology Review, 7, 49–56.
Clayton, M., Sager, R., & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. In European meetings in ethnomusicology, 11, 1–82.
Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians’ body movements. Music Perception: An Interdisciplinary Journal, 24, 433–454.
Danz, A. D., & Janyan, A. (2009). Detecting audio-video tempo discrepancies between conductor and orchestra. In Proceedings of the 31st Annual Conference of the Cognitive Society (pp. 3064–3069).
D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., . . . Fadiga, L. (2012) Leadership in orchestra emerges from the causal relationships of movement kinematics. PLoS One, 7, e35757.
Davidson, J. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21*, 103–113.

Dean, R. T., Byron, T., & Bailes, F. A. (2009). The pulse of symmetry: On the possible co-evolution of rhythm in music and dance. *Musicae Scientiae, 13*, 341–367. doi:10.1177/1029864909013002151

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54*, 277–286.

Engel, A. K., Roelfsema, P. R., Fries, P., Brecht, M., & Singer, W. (1997). Role of the temporal domain for response selection and perceptual binding. *Cerebral Cortex (New York, NY: 1991)*, 7, 571–582.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*, 719–725.

Gilbert, A. (2012, April 6). Connecting music and gesture. *The New York Times*, online. Retrieved from http://archive.nytimes.com/www.nytimes.com/interactive/2012/04/06/arts/music/the-connection-between-gesture-and-music.html

Glowsinski, D., Mancini, M., Cowie, R., Camurri, A., Chiurri, C., & Doherty, C. (2019). The movements made by performers in a skilled quartet: A distinctive pattern, and the function that it serves. *Frontiers in Psychology, 4*, 841. doi:10.3389/fpsyg.2013.00841

Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception, 26*, 427–438.

Green, E. A., & Gibson, M. (2004). *The modern conductor* (7th ed.). Upper Saddle River, NJ: Pearson.

Hilt, P. M., Badino, L., D’Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., & Camurri, A. (2013). Multi-layer adaption of group coordination in musical ensembles. *Scientific Reports, 1*, 5854. doi:10.1038/s41598-019-42395-4

Hove, M. J., & Risen, J. L. (2009). It’s all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27*, 949–960.

Johnson, E. (2014, August 28). Why do so many orchestras lag behind the beat? *The Strad*. Retrieved from http://www.thestrad.com/cpt-latests/why-do-so-many-orchestras-lag-behind-the-beat/

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459–491. doi:10.1037/0033-295X.96.3.459

Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile–visual asynchronous stimuli. *Neuroscience Letters, 430*, 130–134. doi:10.1016/j.neulet.2007.10.044

Kumar, A. B., & Morrison, S. J. (2016). The conductor as visual guide: Gesture and perception of musical content. *Frontiers in Psychology, 7*, Article 1049. doi:10.3389/fpsyg.2013.01049

Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), *Psychology of Time* (pp. 189–232). Bingley, England: Emerald Group Publishing Limited.
Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition, 105*, 533–546. doi:10.1016/j.cognition.2006.11.006

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception, 30*, 71–83.

Price, H. E., & Chang, C. (2005). Conductor and ensemble performance expressivity and state festival ratings. *Journal of Research in Music Education, 53*, 66–77. doi:10.1177/002242940505400106

Price, H. E., & Mann, A. (2011). The effect of conductors on ensemble evaluations. *Bulletin of the Council for Research in Music Education, 189*, 57–72. Retrieved from http://www.jstor.org/stable/10.5406/bulcuoresmedu.189.0057

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969–992. doi:10.3758/BF03206433

Roberts, M. S. (2017, October 9). Why do conductors always conduct ahead of the beat? Classic FM, online. Retrieved from https://www.classicfm.com/discover-music/why-conductors-beat-ahead/

Rudolf, M. (1980). *The grammar of conducting*. New York, NY: Schirmer.

Schiavio, A., & Høffding, S. (2015). Playing together without communicating? A pre-reflective and inactive account of joint musical performance. *Musicae Scientiae, 19*, 366–388. doi:10.1177/1029864915593333

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36, 868–897. doi:10.1068/p5635

Silvey, B. A., Wacker, A. T., & Felder, L. (2017). Effects of baton usage on college musicians’ perceptions of ensemble performance. *International Journal of Music Education, 35*, 333–342. doi:10.1177/0255761416667465

Sparenberg, P., Topolinski, S., Springer, A., & Prinz, W. (2012). Minimal mimicry: Mere effector matching induces preference. *Brain and Cognition, 80*, 291–300. doi:10.1016/j.bandc.2012.08.004

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*, 971–995. doi:10.3758/s13414-010-0073-7

Timmers, R., Endo, S., Bradbury, A., & Wing, A. M. (2014). Synchronization and leadership in string quartet performance: A case study of auditory and visual cues. *Frontiers in Psychology, 5*, 645.

Trost, W. J., Labbé, C., & Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia, 96*, 96–110. doi:10.1016/j.neuropsychologia.2017.01.004

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 14580–14585. doi:10.1073/pnas.1221454110

Tsay, C. J. (2014). The vision heuristic: Judging music ensemble by sight alone. *Organizational Behavior and Human Decision Processes, 124*, 24–33. doi:10.1016/j.obhdp.2013.10.003

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research, 1111*, 134–142. doi:10.1016/j.brainres.2006.05.078

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2006). Dimensions of emotion in expressive music performance. *Annals of the New York Academy of Sciences, 1060*, 462–466. doi:10.1196/annals.1360.052