Sight over sound in the judgment of music performance

Chia-Jung Tsay[1]

Department of Management Science and Innovation, Faculty of Engineering Science, University College London, London WC1E 6BT, United Kingdom

Edited by Dale Purves, Duke-National University of Singapore Graduate Medical School, Singapore, Singapore, and approved July 12, 2013 (received for review December 21, 2012)

Social judgments are made on the basis of both visual and auditory information, with consequential implications for our decisions. To examine the impact of visual information on expert judgment and its predictive validity for performance outcomes, this set of seven experiments in the domain of music offers a conservative test of the relative influence of vision versus audition. People consistently report that sound is the most important source of information in evaluating performance in music. However, the findings demonstrate that people actually depend primarily on visual information when making judgments about music performance. People reliably select the actual winners of live music competitions based on silent video recordings, but neither musical novices nor professional musicians were able to identify the winners based on sound recordings or recordings with both video and sound. The results highlight our natural, automatic, and nonconscious dependence on visual cues. The dominance of visual information emerges to the degree that it is overweighted relative to auditory information, even when sound is consciously valued as the core domain content.

social perception | cognition | decision making | evaluation | communication

We do judge books by their covers. We prefer the nicely wrapped holiday gifts (1), fall in love at first sight (2), and vote for the politician who looks most competent (3). Daily life is littered with examples of how visual information can have a powerful effect on social cognition, ranging from interpersonal perception to consumer judgment (4–7).

In music, however, it is auditory information that defines the domain. Hiring committees have embraced "blind" screenings (8) not only out of the pursuit of fairness, but also in response to critics who disparage those who prioritize visually stimulating choreography over the composer's intended sound (9, 10). Professional musicians consistently report that sound is the most important information in the evaluation of music (11). After all, the foundation of the field was built upon the creation of a better sound; ear-training classes are part of the core curriculum at major conservatories, and performance is evaluated during auditions.

Given the wide consensus that sound is central to judgment about performance in music (12), our judgments should be limited if we are denied access to sound. Although people often make evaluations quickly on the basis of visual cues (4–7, 13, 14), these cues have traditionally been neglected (15) and discounted as peripheral to the meaning of music (16). However, people can lack insight into their own preferences and cognitive processes (17–19), or be unable or unwilling to report their beliefs (20, 21). These findings suggest that there may be gaps between what we say we use to evaluate performance and what we actually use. People may be unlikely to recognize or admit that visual displays can affect their judgment about music performance, particularly in a domain in which other signals are deemed to be more indicative of quality.

Using real competition outcomes, this series of experiments empirically tests the impact of visual information on expert judgment. In highly competitive arenas such as music, competitions emerge as one launching pad for establishing careers. With these important decisions at stake, professionals are sought for their expertise to identify the best. Indeed, no matter what domain, the judgment of performance occupies a key area of investment. Experts are trained and societal institutions are constructed to identify, develop, and reward the highest levels of achievement. We trust that professionals can judge performance through their specialized knowledge; these are the leaders who are responsible for shaping the landscape of the future of their fields. In music, we expect that professionals would critique the sound of music.

However, research points to the influence of visual information on the perception and processing of sound (22, 23), extending even to the domain of music (16, 24). Given that the literature suggests that either audition (25–27) or vision (28–30) may dominate, and that the two modalities can be complementary (31–35) and share many similarities in their cognitive processing (36, 37), these experiments offer a direct comparison of the extent to which auditory versus visual cues affect our evaluations and decision making. It may be that, regardless of training, knowledge, and theories about the meaning of music, experts are just as vulnerable as novices to certain heuristics—ones that may be at odds with what is valued by the field.

Honing in more specifically on the music psychology literature, there has been great interest in investigating performance evaluation and expert evaluators with more precision (38). As a host of factors that contribute to performance assessment have not been well understood or considered (19), a fuller understanding of the evaluation process holds great promise. The role that auditory versus visual information plays in performance evaluation is of particular interest to researchers, practitioners, and educators. It thus becomes more surprising that, with some exception (39), there has been relatively insufficient empirical research to justify definitive conclusions (38). An understanding that is grounded in empirical research lends itself not only to the possibility of more objective evaluation processes, but also to the crafting of more effective performance.

With the general consensus on the importance of sound in the domain of music, as "an art of sound" (40), it follows that experts and key decision makers would privilege auditory-related rating in professional evaluation and assessment, even when such items show insufficient reliability (41–45). However, despite all that is invested in the auditory domain, low interrater correlations suggest that such basis of evaluation is an unreliable process. The increasing interest in investigations of the role of visual information in evaluation (24, 39) dovetails well with recent calls for the need to include the visual component in music performance

Author contributions : C.T. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]E-mail: c.tsay@ucl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221454110/-/DCSupplemental.

(46) and the authenticity that this modality specifically communicates through expressive behavior (47).

The current research uses a two-pronged approach: (*i*) the experimental design offers high test power and tight control over variables of interest, allowing for better substantiated conclusions, and (*ii*) the use of field data with real decision processes and outcomes addresses external validity and relevance for a broad range of contexts that involve performance evaluation. Given the questionable reliability of expert ratings based on audio-only information, and the recent works demonstrating the substantial role of visual information (8, 22, 24), it may be that a visual dominance would emerge above and beyond the impact of auditory information.

In this set of experiments, participant responses were used to extrapolate the evaluation processes of the original expert judges and determine which cues—visual or auditory—were most influential for their decisions in arriving at the real-time results of live music competitions. Given different versions of competition performances, 1,164 participants in total were asked to identify the actual competition winners. These choices were then compared against the established outcomes, previously decided by panels of expert judges (*SI Text*). As a domain in which sound is central to what experts and novices alike value about performance, music offers a strong test of the impact of visual information on the judgment of performance.

Results

Experiment 1: Core Beliefs About Music. Suppose that you have the chance to win cash bonuses if you can guess who won a live music competition. You may choose the type of recording you think would give you the best chance at winning the prize. You can select sound recordings, video recordings, or recordings with both video and sound. Which recordings do you choose? In experiment 1, participants were asked to make exactly that decision and bet their study earnings on their choices.

As expected, 58.5% chose the sound recordings, significantly more so than the 14.2% who chose video recordings, $\chi^2(1, n = 77) = 28.89$, $P < 0.001$. Despite a "tax" levied on selecting the recordings with both video and sound, 27.4% still chose those recordings, a significantly larger proportion than those who chose the video recordings, $\chi^2(1, n = 44) = 4.46$, $P = 0.035$. People have the intuition that sound is a more revealing channel of information in the domain of music and that recordings with both visual and auditory output offer additional and more relevant information that better approximates the conditions under which the original expert decisions were made (*SI Text*).

Experiments 2–5. In experiments 2–5, the top three finalists in each of 10 prestigious international classical music competitions were presented to participants. Given such difficult decisions (*SI Text*), untrained participants should fare no better than chance (33%) in identifying the winners of these competitions. In fact, even expert interrater agreement tends to be moderate, hovering at an average of 67%; consensus is notoriously absent (48).

Novice participants. In experiment 2, novice participants were presented with both video-only and sound-only versions of 6-s clips of the top performances from international competitions. Although 83.3% of participants reported that the sound mattered most for their evaluation of music performance, these same participants were significantly more likely to identify the winners when they were presented with only the visual components of the performances, $t_1(105) = 12.07$, $P < 0.001$; Cohen's d = 1.18 (Fig. 1). The item analysis indicated that the effect held across all 10 competitions, $t_2(9) = 4.37$, $P = 0.002$. Indeed, with silent video-only recordings, participants were significantly above chance (52.5%), $t(105) = 10.90$, $P < 0.001$. With sound-only recordings, they were significantly below chance (25.5%) at identifying the winners, $t(105) = -5.23$, $P < 0.001$.

As seen in experiment 1, participants believed that recordings with both video and sound would allow them to best approximate the original expert judgments. Is it the case that more information necessarily leads to better judgment? Experiment 3 tested judgment when more information was available, and presented participants with video-only, sound-only, or video-plus-sound versions of the performance clips included in experiment 2. Participants performed below chance with sound-only recordings (28.8%), $t(66) = -2.09$, $P = 0.040$, and at chance with video-plus-sound recordings (35.4%), $t(67) = 0.94$, P = not significant (n.s.). However, with silent video-only recordings, 46.4% of novices were able to identify the winners, $t(49) = 4.04$, $P < 0.001$.

These findings suggest that novices are able to approximate expert judgments, originally made after hours of live performances, with brief, silent video recordings. However, when novices were also given the sound of the performances through the video-plus-sound recordings, they did no better than picking a winner at random (*SI Text*). As surprising as these findings are, they may be

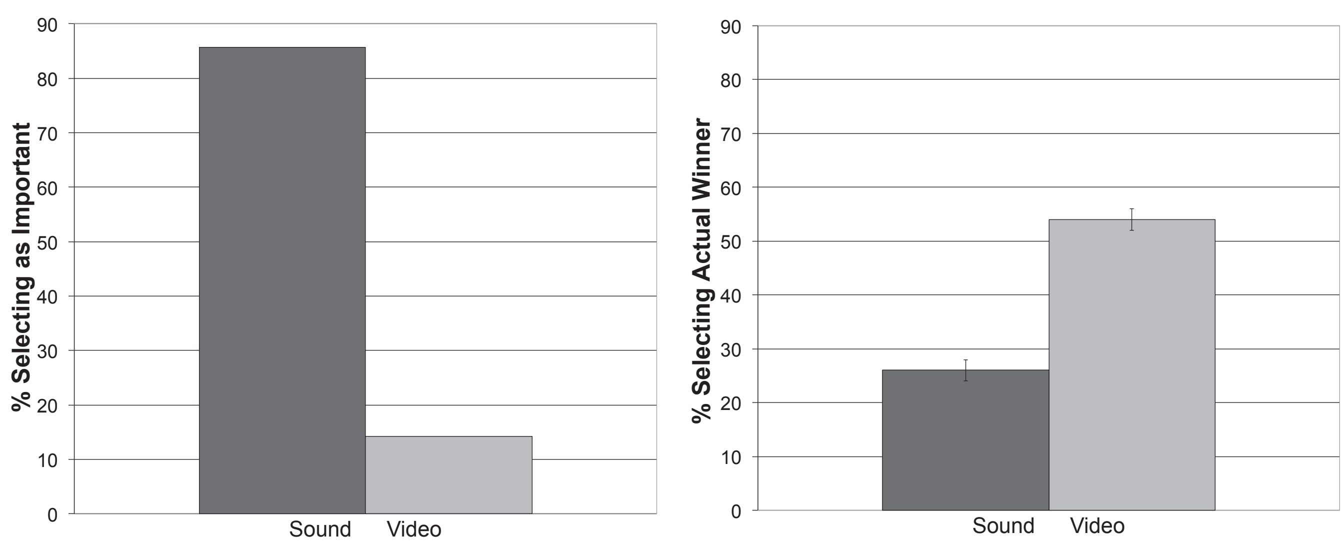

Fig. 1. A comparison of the reported importance of sound vs. visuals for evaluation (*Left*), with the % novices identifying actual competition outcomes when given sound-only vs. video-only stimuli (*Right*), in experiment 2 ($n = 106$).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

due to novices' lack of music training, which forces them to rely on visual cues.

Expert participants. Using the same sets of competition clips and paralleling the design in experiments 2 and 3, experiments 4 and 5 explored whether the dominance of visual cues remains in domain experts. Professional musicians have the knowledge and training to discern the quality of performance through sound; they should be able to outperform novices in identifying the actual winners. Although the assumed superior judgment of experts is dependent on domain and context (49, 50), these musicians had participated in and judged competitions and are familiar with how professional judgment is determined.

In experiment 4, 96.3% of domain-expert participants reported that the sound mattered more for their evaluations, $\chi^2(1, n = 27) = 23.15$, $P < 0.001$. Despite musicians' training to use and value sound in their evaluations, only 20.5% of experts identified the winners when they heard sound-only versions of the recordings, $t(34) = -6.11$, $P < 0.001$. However, 46.6% did so upon viewing silent video clips, $t(34) = 4.05$, $P < 0.001$. Those with video-only stimuli performed significantly better, compared with those who heard sound-only stimuli, $t_1(34) = 5.89$, $P < 0.001$; Cohen's d = 1.01 (Fig. S1). An item analysis indicates that this effect held across all 10 competitions, $t_2(9) = 3.74$, $P = 0.005$.

In experiment 5, 82.3% of professional musicians cited sound as the most important information for judgment, $\chi^2(2, n = 96) = 103.56$, $P < 0.001$. However, when provided sound, only 25.7% of experts were able to identify the actual winners (Fig. 2), a rate worse than chance, $t(29) = -3.34$, $P = 0.002$. With video-only stimuli, musicians performed significantly better than chance (47.0%) at identifying the actual winners, $t(32) = 3.40$, $P = 0.002$. Experts were significantly better with video-only stimuli than with sound-only stimuli, $t_1(61) = 4.48$, $P < 0.001$; Cohen's d = 1.20. An item analysis indicates that these effects were robust across all 10 competitions, $t_2(9) = -2.36$, $P = 0.04$.

In the third condition in this experiment, when provided with stimuli with both video and sound, experts were again at chance (*SI Text*) at 29.5%, $t(39) = -1.43$, P = n.s. They were not significantly better than those who received sound-only stimuli, $t(48) = 1.33$, P = n.s. Those who received video-only stimuli, even compared with those who received both video and sound, were still significantly more likely to approach the actual outcomes, $t(71) = 3.72$, $P < 0.001$.

Experts were not significantly different from novices in their judgments of music performance. Novices and experts are similarly below chance with sound recordings and at chance with recordings with both video and sound. Novices and experts also paralleled each other in their use of different cues to arrive at the competition outcomes made by the original judges, with no significant differences through the sound-only recordings, $t(95) = 0.85$, P = n.s.; the video-plus-sound recordings, $t(106) = 1.68$, P = n.s.; nor the video-only recordings, $t(81) = -0.12$, P = n.s.

In supplemental tests of the primacy of visual cues, additional studies featuring the same between-subjects design as experiments 3 and 5 replicate the findings outlined in this paper with 3-s and 1-s recordings. The at-chance findings with sound-only and video-plus-sound recordings remain even with longer time intervals ranging up to 60-s recordings. These results suggest that the findings outlined in the current experiments remain meaningful for more extended periods of evaluation.

These results demonstrate how visual information, the information generally deemed as peripheral in the domain of music, can be overweighted when such inclination is neither valued nor recognized. Ironically, this tendency results in our neglect of the most relevant information: the sound of music. What then are novices and experts paying attention to when making their judgments? The next two experiments examine the mechanisms that account for the primacy of visual cues and our dependence on visual information. The studies explore the types of visual information that are used in judgment and how motion, emotion, and apparent motivation contribute to professional inferences about the quality of music performance (*SI Text*).

Experiments 6 and 7: Mechanism. Movement and gesture are elements of performance that are primarily visual. Experiment 6 examined whether motion impacts the professional judgment of music performance. In this study, recordings were distilled to their most basic representation as outlines of motion (Fig. S2). After seeing these 6-s silent clips of the three finalists, participants were asked to identify the actual winners. Participants were significantly better than chance (48.8%) at identifying the outcomes, $t(88) = 6.49$, $P < 0.001$. Viewing brief motion alone allowed an approximation of professional judgment made after hours of live performance with both visual and auditory information.

The importance of dynamic visual information to professional judgment was further established through two supplementary experiments (*SI Text*). Although demographic cues such as race and sex have been associated with various capabilities (51, 52), such as the quality of musicianship (8)—and although the many advantages of physical attractiveness have been documented (53), from hiring (54) to income (55)—these static visual cues did not significantly impact professional judgment in these competitions.

Visual information may be powerful through its associations with expressive behavior (16, 56) and through its emotional impact. Professional musicians may value novelty (57), involvement

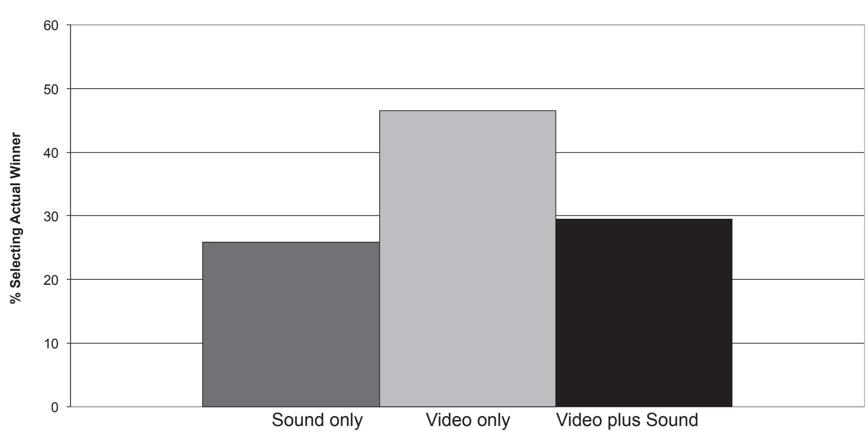

Fig. 2. The % professional musicians identifying actual competition outcomes given sound-only, video-only, or video-plus-sound stimuli, in experiment 5 ($n = 103$). Thirty-three percent indicates an identification rate at chance.

(58), motivation, and passion (59) as essential to the quality of creative performance. These attributes may be more visible than they are audible. Furthermore, observers not only may perceive nonverbal cues, but also may experience more intense emotional changes and foster greater interpersonal understanding through these nonverbal cues through emotional contagion (60, 61). In the domain of music, however, sound is often assumed to be the primary medium through which creative and affective expression is conveyed and understood (62, 63).

In experiment 7, 262 participants were presented with either video-only or sound-only 6-s recordings of the competition performances. They were then asked to identify the most confident, creative, involved, motivated, passionate, and unique performer in each set of three finalists in the competitions. These evaluations were then compared against the original competition outcomes. Creativity, involvement, motivation, passion, and uniqueness were significantly more salient through visual cues rather than through sound.

Passion had considerable impact on the professional judgment of quality when it was visible; through silent videos, those selecting "the most passionate contestant" identified the actual winners at rates significantly higher than chance (59.6%). They also fared better than those making the same judgments through audio recordings (38.7%), $t(196) = 7.01$, $P < 0.001$. Involvement (53.1%), motivation (52.8%), creativity (44.6%), and uniqueness (43.6%) also contributed to the visual information that signaled quality of performance in a way that auditory information did not allow either novice or expert participants to perceive (all P's < 0.001). Confidence was not a factor that allowed participants to distinguish among the performers through either visuals or sound, $t(193) = -0.68$, P = n.s.

The final experiments explored the visual elements that contribute to the professional judgment of music. Motion, motivation, creativity, and passion are perceived as hallmarks of great performance (*SI Text*). As those facets of performance are visually accessible and readily so, they may be universally understood throughout levels of expertise. Thus, even novices are able to quickly identify the actual winners among world-class performers, without being encumbered by the sound of music that professional musicians unintentionally and nonconsciously discard.

These additional experiments suggest that performers' movements may contribute substantially toward inferences about the quality of performance. Our movements facilitate aspects of cognitive abilities (64, 65) such as coordination and the appreciation of rhythm (66). The sight of others' gestures may also influence our understanding about music. Our responsiveness to movement (67–69) and emotional expression (62, 63, 70) may underlie the intuition that musicians' motions and emotions represent exceptional performance. Future work will be needed to test not only our perceptions of performers, but also the emotions evoked in audiences, to better understand the affective contributions to the primacy of visual cues in the judgment of performance.

Discussion

This set of seven experiments (Table S1) suggests that novices' judgment mirrors that of professionals; both novices and experts make judgments about music performance quickly and automatically on the basis of visual information. Given the relative lack of consensus about competition outcomes noted among even expert judges, the fact that novices are able to quickly identify the actual competition winners at such high rates through silent videos alone is of both statistical and practical significance. These findings point to a powerful effect of vision-biased preferences on selection processes even at the highest levels of performance.

Experts and novices alike privilege visuals above sound, the very information that is explicitly valued and reported as core to decision making in the domain of music. Moreover, when sound is made available along with the video, it led people away from the actual (visually based) competition outcomes. This finding complements those of a recent landmark meta-analysis, which argues for an influence of the visual component on music performance evaluation in a multiplicative cross-modal model of perception (24). When both sound and visuals were available in the current work, judgments appear to be impacted by both modalities.

Ongoing research suggests that pressures that constrain our cognitive resources may lead to a visual dependence. As the current work focuses on choices made during competitive settings, more information would not necessarily lead to better approximations of expert judgment, even if it increases confidence in judgment (71). People are limited by attention to certain cues, with inconsistency (72, 73) and at times detriments to judgment (74).

Professional musicians and competition judges consciously value sound as central to this domain of performance, yet they arrive at different winners depending on whether visual information is available or not. This finding suggests that visual cues are indeed persuasive and sway judges away from recognizing the best performance that they themselves have, by consensus, defined as dependent on sound. Professional judgment appears to be made with little conscious awareness that visual cues factor so heavily into preferences and decisions.

Both musical novices and professional musicians reported attempting to identify the highest quality performances. These self-reports are further supported by the studies that implemented incentives and bonuses for participant performance in identifying the actual winners. However, both experts and novices appear to be surprised by their own data, and experts in particular reported a severe lack of confidence in their judgment when they were assigned to the video-only recordings, not knowing that their approximations of the actual outcomes would be superior under such constrained conditions. The notion that our experience of music (75) depends so much on visual information—at a nonconscious level and to a degree that interferes with what people actually value—points to consequential implications (*SI Text*).

Against broad consensus that auditory information is core to the domain of music, these experiments offer strong tests of the primacy of visual information. The implications of these findings thus extend to any context that calls for the professional judgment of performance. Ongoing research suggests that the effects are generalizable to multiple domains, such as management and entrepreneurship—as well as to multiple levels, from individuals to groups.

The dominance of visual information in our decision circuitry may have evolved as adaptive (76, 77) and reliable, evocative of how visual circuitry itself is molded by accumulated experience and successfully guided behavior (78, 79). However, when these decisions involve other information more predictive of performance, whether it concerns hiring employees, interviewing physicians, or selecting political leaders, we must be more mindful of our inclination to depend on visual information at the expense of the content that we actually value as more relevant to our decisions. Given the dominance of visual cues in our decision making, it would be valuable to determine the contexts in which a visual dependence may not be one that leads to wise decisions and good long-term investments in selecting, promoting, and rewarding talent.

Professional training may hone musicians' technical prowess and cultivate their expressive range, but in this last bastion of the realm of sound, it does little to shift our natural and automatic overweighting of visual cues. After all, sound can be neglected while trained "ears" focus on the more salient visual cues. It is unsettling to find—and for musicians not to know—that they themselves relegate the sound of music to the role of noise.

Materials and Methods

The Harvard University Institutional Review Board approved all procedures. Informed consent was obtained from all participants.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Experiment 1. One hundred six participants ($M_{age} = 20.73$, SD = 2.46; 49.5% male*) volunteered.† Participants were instructed about 10 live classical music competitions that they would judge, based on excerpts of the three finalists in each competition. They had the chance to receive an additional $8 if their selections matched the actual competition outcomes. They had the choice of sound or video recordings; or, if they chose the recordings with both video and sound, $2 would be deducted from any bonuses won.

Experiment 2. One hundred six participants ($M_{age} = 22.26$, SD = 1.79; 41.1% male*) with little to no experience in classical music volunteered.† Through a within-subjects design, each participant received both the video-only set and the sound-only set of the same performances (*SI Text*). Participants were then asked to identify the winner of each competition. Finally, they were asked to identify whether sound, visuals, or other cues were more important for them in judging a music competition.

Experiment 3. One hundred eighty-five participants ($M_{age} = 24.18$, SD = 9.64; 46.1% male*) with little to no experience in classical music volunteered.† Through a between-subjects design, participants were randomly assigned to one of three conditions: video-only, sound-only, or video-plus-sound versions of the experiment 2 stimuli. They were then asked to identify the winners and report whether sound, visuals, or other cues were more important for them in judging a music competition (*SI Text*).

Experiment 4. Thirty-five professional musicians ($M_{age} = 27.00$, SD = 9.69; 31.6% male) volunteered. They were recruited from music conservatories, symphony orchestras, and professional music organizations. The design paralleled the within-subjects format used in experiment 2 and implemented the same stimuli (*SI Text*).

Experiment 5. One hundred six professional musicians ($M_{age} = 27.25$, SD = 12.55; 41.5% male) volunteered. The design paralleled the between-subjects format used in experiment 3 and implemented the same stimuli. Analyses on effects of demographic variables revealed no significant patterns (*SI Text*).

Experiment 6. Eighty-nine participants ($M_{age} = 27.38$, SD = 10.68; 50.0% male*) volunteered.† Participants received silent videos from the experiment 2–5 stimuli that had been reduced to black-and-white moving outlines (Fig. S2). Participants were then asked to identify the winners of each competition.

Experiment 7. Two hundred sixty-two participants ($M_{age} = 21.52$, SD = 3.36; 52.3% male) volunteered.† Participants were assigned to either the silent videos or the audio recordings from the experiment 2–5 stimuli. They were then asked to identify the most confident, creative, involved, motivated, passionate, and unique performer in each set of finalists. Repeat choices were allowed.

*Participants who did not report their sex were not included in the calculation.

†Participants were recruited from a community sample in the northeastern United States and were paid $20 for their participation in an hour-long set of unrelated studies that included the current experiment.

ACKNOWLEDGMENTS. I thank M. Bazerman for his insights and invaluable support throughout the development of this work; T. Amabile, S. Barsade, and B. Groysberg for generous feedback and guidance; N. Ambady and K. Nakayama for the forum that sparked the initial studies; J. Abraham, R. Cheong, S. Dobrow, P. Fili, A. Galinsky, J. Golann, M. Morgan, O. Obodaru, D. Parno, K. Shonk, and the University College London MSI reading group for input on earlier drafts; and M. Banaji, K. Hida, R. Jean, J. Lee, J. Polzer, C. Silva, B. Simpson, B. Tang, A. Williams, D. Wollenstein, and members of non-lab and the HBS OB lab for research help. Many thanks also to the dedicated musicians who shared their art and minds. This research was supported by Harvard University and the Wyss Foundation.

1. Howard DJ (1992) Gift-wrapping effects on product attitudes: A mood-biasing explanation. *J Consum Psychol* 1(3):197–223.
2. Back MD, Schmukle SC, Egloff B (2010) Why are narcissists so charming at first sight? Decoding the narcissism-popularity link at zero acquaintance. *J Pers Soc Psychol* 98(1): 132–145.
3. Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence from faces predict election outcomes. *Science* 308(5728):1623–1626.
4. Ambady N, Rosenthal R (1993) Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J Pers Soc Psychol* 64(3): 431–441.
5. Ambady N, Rosenthal R (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol Bull* 111:256–274.
6. Ballew CC, Todorov A (2007) Predicting political elections from rapid and unreflective face judgments. *Proc Natl Acad Sci USA* 104(46):17948–17953.
7. Platz F, Kopiez R (2013) When the first impression counts: Music performers, audience, and the evaluation of stage entrance behavior. *Music Sci* 17(2):167–197.
8. Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. *Am Econ Rev* 90:715–741.
9. Tommasini A (November 10, 2003) A showman revs up the classical genre. *NY Times*. Available at www.nytimes.com/2003/11/10/arts/music-review-a-showman-revs-up-the-classical-genre.html.
10. von Rhein J (February 9, 2002) Restraint not part of Lang's repertoire. *Chicago Tribune*. Available at http://articles.chicagotribune.com/2002-02-09/news/0202090304_1_lang-lang-lower-brass-tchaikovsky-fifth-symphony.
11. Murnighan K, Conlon D (1991) The dynamics of intense work groups: A study of British string quartets. *Adm Sci Q* 36(2):165–186.
12. Sloboda JA, Lamont A, Greasley AE (2008) *The Oxford Handbook of Music Psychology*, eds Hallam S, Cross I, Thaut M (Oxford Univ Press, Oxford), pp 431–440.
13. Benjamin DJ, Shapiro JM (2009) Thin-slice forecasts of gubernatorial elections. *Rev Econ Stat* 91(33):523–536.
14. Rule NO, Ambady N (2008) The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychol Sci* 19(2):109–111.
15. Gabrielsson A (2003) Music performance research at the millennium. *Psychol Music* 31(3):221–272.
16. Bergeron V, Lopes DM (2009) Hearing and seeing musical expression. *Philos Phenomenol Res* 78(1):1–16.
17. Burson KA, Larrick RP, Klayman J (2006) Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *J Pers Soc Psychol* 90(1):60–77.
18. Nisbett RE, Wilson TD (1977) Telling more than we can know: Verbal reports on mental processes. *Psychol Rev* 84:231–259.
19. McPherson GE, Schubert E (2004) *Musical Excellence: Strategies and Techniques to Enhance Performance*, ed Williamon A (Oxford Univ Press, New York), pp 61–82.
20. Greenwald AG, Banaji MR (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol Rev* 102(1):4–27.
21. Schlenker B, Dlugolecki D, Doherty K (1994) The impact of self-presentations on self-appraisals and behaviors: The power of public commitment. *Pers Soc Psychol Bull* 20: 20–33.
22. Thompson WF, Russo FA (2007) Facing the music. *Psychol Sci* 18(9):756–757.
23. Sams M, et al. (1991) Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 127(1):141–145.
24. Platz F, Kopiez R (2012) When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept* 30(1): 71–83.
25. Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. *Nature* 385(6614):308.
26. Shams L, Kamitani Y, Shimojo S (2000) What you see is what you hear: Sound induced visual flashing. *Nature* 408:788.
27. Shipley T (1964) Auditory flutter-driving of visual flicker. *Science* 145(3638):1328–1330.
28. Bertelson P, Aschersleben G (1998) Automatic visual bias of perceived auditory location. *Psychon Bull Rev* 5:482–489.
29. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264(5588): 746–748.
30. Botvinick M, Cohen J (1998) Rubber hands 'feel' touch that eyes see. *Nature* 391(6669):756.
31. Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10(9):873–888.
32. Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11(12):535–543.
33. Collignon O, et al. (2008) Audio-visual integration of emotion expression. *Brain Res* 1242:126–135.
34. de Gelder B, Böcker KB, Tuomainen J, Hensen M, Vroomen J (1999) The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neurosci Lett* 260(2):133–136.
35. de Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. *Trends Cogn Sci* 7(10):460–467.
36. Barnett-Cowan M, Harris LR (2009) Perceived timing of vestibular stimulation relative to touch, light and sound. *Exp Brain Res* 198(2-3):221–231.
37. Harrar V, Harris LR (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp Brain Res* 186(4): 517–524.
38. Finnäs L (2001) Presenting music live, audio-visually or aurally—does it affect listeners' experiences differently? *Br J Music Educ* 18(1):55–78.
39. Davidson JW, Malloch S (2009) *Communicative Musicality: Exploring the Basis of Human Companionship*, eds Malloch S, Trevarthen C (Oxford Univ Press, New York), pp 565–583.
40. Davies S (1994) *Musical Meaning and Expression* (Cornell Univ. Press, Ithaca NY).
41. Fiske HE (1975) Judge-group differences in the rating of secondary school trumpet performances. *J Res Music Educ* 23(3):186–196.
42. Fiske HE (1977) Relationship of selected factors in trumpet performance adjudication reliability. *J Res Music Educ* 25(4):256.

43. Fiske HE (1983) Judging musical performances: Method or madness? *J Res Music Educ* 1(1):7–10.
44. Bergee MJ (2003) Faculty interjudge reliability of music performance evaluation. *J Res Music Educ* 51(2):137.
45. Thompson S, Williamon A (2003) Evaluating evaluation: Musical performance assessment as a research tool. *Music Percept* 21(1):21–41.
46. Thompson S (2007) Determinants of listeners' enjoyment of a performance. *Psychol Music* 35(1):20–36.
47. Auslander P (2008) *Liveness. Performance in a Mediatized Culture* (Routledge, New York).
48. Schonberg H (July 6, 1986) Do today's pianists have the romantic touch? *NY Times*. Available at www.nytimes.com/1986/07/06/arts/do-today-s-pianists-have-the-romantic-touch.html.
49. Castel AD, McCabe DP, Roediger HL, Heitman JL (2007) The dark side of expertise: Domain-specific memory errors. *Psychol Sci* 18(1):3–5.
50. Englich B, Mussweiler T, Strack F (2006) Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Pers Soc Psychol Bull* 32(2):188–200.
51. Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment in labor market discrimination. *Am Econ Rev* 94(4): 991–1013.
52. Pager D, Western B, Bonikowski B (2009) Discrimination in low-wage labor markets. *Am Sociol Rev* 74(5):777–799.
53. Dion KK, Berscheid E, Walster E (1972) What is beautiful is good. *J Pers Soc Psychol* 24(3):285–290.
54. Beehr TA, Gilmore DC (1982) Applicant attractiveness as a perceived job relevant variable in selection. *Acad Manage J* 25:607–617.
55. Biddle J, Hamermesh D (1998) Beauty, productivity, and discrimination: Lawyers' looks and lucre. *J Labor Econ* 16(1):172–201.
56. Allport GW (1937) *Personality: A Psychological Interpretation* (Henry Holt, New York).
57. Amabile TM (1983) *The Social Psychology Of Creativity* (Springer, New York).
58. Ruscio J, Whitney D, Amabile TM (1998) Looking inside the fishbowl of creativity: Verbal and behavioral predictors of creative performance. *Creat Res J* 11:243–263.
59. Amabile TM, Kramer SJ (2007) Inner work life: Understanding the subtext of business performance. *Harv Bus Rev* 85(5):72–83, 144.
60. Barsade S (2002) The ripple effect: Emotional contagion and its influence on group behavior. *Adm Sci Q* 47:644–675.
61. Nummenmaa L, et al. (2012) Emotions promote social interaction by synchronizing brain activity across individuals. *Proc Natl Acad Sci USA* 109(24):9599–9604.
62. Trainor L, Schmidt L (2003) *The Cognitive Neuroscience of Music*, eds Peretz I, Zatorre R (Oxford Univ Press, Oxford), pp 310–324.
63. Trehub SE (2003) The developmental origins of musicality. *Nat Neurosci* 6(7):669–673.
64. Cook SW, Mitchell Z, Goldin-Meadow S (2008) Gesturing makes learning last. *Cognition* 106(2):1047–1058.
65. Dijkstra K, Kaschak MP, Zwaan RA (2007) Body posture facilitates retrieval of autobiographical memories. *Cognition* 102(1):139–149.
66. Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308(5727):1430.
67. Wöllner C, Deconinck FJA, Parkinson J, Hove MJ, Keller PE (2012) The perception of prototypical motion: Synchronization is enhanced with quantitatively morphed gestures of musical conductors. *J Exp Psychol Hum Percept Perform* 38(6):1390–1403.
68. Keller PE, Knoblich G, Repp BH (2007) Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Conscious Cogn* 16(1):102–111.
69. Sevdalis V, Keller PE (2011) Captured by motion: Dance, action understanding, and social cognition. *Brain Cogn* 77(2):231–236.
70. Van Zijl A, Sloboda JA (2010) Performers' experienced emotions in the construction of expressive musical performance: An exploratory investigation. *Psychol Music* 39(2): 196–219.
71. Tsai C, Klayman J, Hastie R (2008) Effects of amount of information on judgment accuracy and confidence. *Organ Behav Hum Decis Process* 107:97–105.
72. Chinander K, Schweitzer M (2003) The input bias: The misuse of input information in judgments of outcomes. *Organ Behav Hum Decis Process* 91:243–253.
73. Hsee C (1996) The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organ Behav Hum Decis Process* 67:247–257.
74. Olivola C, Todorov A (2010) Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *J Exp Soc Psychol* 46:315–324.
75. Sloboda J (2008) Science and music: The ear of the beholder. *Nature* 454(7200):32–33.
76. Jerison HJ (1973) *Evolution of the Brain and Intelligence* (Academic, New York).
77. Jurmain R, Kilgore L, Trevathan W (2008) *Essentials of Physical Anthropology* (Wadsworth, Belmont, CA).
78. Purves D, Lotto RB (2011) *Why We See What We Do Redux: A Wholly Empirical Theory of Vision* (Sinauer, Sunderland, MA).
79. Purves D, Wojtach WT, Lotto RB (2011) Understanding vision in wholly empirical terms. *Proc Natl Acad Sci USA* 108(3, Suppl 3):15588–15595.

Supporting Information

Tsay 10.1073/pnas.1221454110

SI Text

A conservative test is one that is less likely to find statistical significance. Given the counterintuitive findings in this work, it was important that findings are well-replicated, effects are robust, and conservative tests be used, so that we might conclude that the results are due to real effects. The use of the domain of music thus offers a conservative test of the impact of visual information on judgment, given that it is the domain in which auditory information should be much more important than visual information.

First, experiment 1 tested our core beliefs about the importance of auditory versus visual output for the judgment of performance. Then, experiments 2–5 explored the relative impact of the two modalities on our judgment. Participants ranged from those with little or no music training (experiments 2 and 3) to professional musicians (experiments 4 and 5). To compare the relative influence of visual versus auditory information, the same participants were presented with both video-only and sound-only recordings in a within-subjects design in experiments 2 and 4 whereas different sets of participants were given video-only, sound-only, or video-plus-sound recordings in a between-subjects design in experiments 3 and 5. Finally, experiments 6 and 7 examined the mechanisms underlying the impact of visual information on judgment.

SI Results

Experiments 2–5. International classical music competitions do not publicize details about individual voting. Given that data regarding confidential procedures are not made public, interviews were conducted with established authorities (judges and competition directors) to gain a better understanding about the competition judging conditions and criteria. As quantitative and empirical data were not available from the international competitions themselves, the rate of interrater agreement was based on an average of estimates provided by expert sources. Further investigation with the organization that oversees these competitions reveals a similar lack of unanimity regarding the choice of the winner.

Robustness. Although it could be argued that visual versus audio stimuli convey different amounts of information in the same brief moments, in this set of experiments, the audio excerpts and video excerpts represent identical musical content and the identical measures of musical compositions. In addition, although the physical movement of the performer and thus the visual content can vary a great deal across a short performance, that same physicality is the means through which the range of audio content is produced and, thus, also perceptible.

Alternative explanations. Decisions based on sound alone were not randomly distributed (all Ps < 0.05, one-sample t tests of rates at which actual winners were chosen, as listed in Table S1)—that is, there is variance in auditory information, and the quality of the music performances was distinguishable and not acoustically equivalent. Sound did allow for differentiation among performers, yet people still surprisingly relied primarily on visual information in their judgment.

Participants who were randomly assigned to receive sound-only recordings were able to choose one performer over the other two performers in each trial. Additional analyses of these decisions indicate that they were significantly different from at-chance choices (all Ps < 0.05, χ^2 tests of frequencies of each of three performers per trial being chosen by participants), which would have resulted had the acoustic quality of the music performances been indistinguishable. However, the winners as chosen by participants who received sound-only recordings were frequently not the actual winners of the live round competitions, as those choices had been impacted more by visual information. This observation points to intriguing future research directions regarding the external validity and predictive power of a visual dominance, such as whether expert judges' visually based evaluations are predictive of short-term and/or long-term success.

Finally, an additional study was conducted with a separate population of 90 expert participants, which further demonstrates that not only is there enough variance in auditory information to allow participants to select one individual over the others at rates over chance, but also that the degree of variance perceived through sound-only excerpts is similar to the degree of variance perceived through video-only excerpts.

In this study, participants were randomly assigned to receive either the sound-only, video-only, or video-plus-sound recordings used in experiments 2–5. They were then asked to evaluate on a 0–100 scale the quality of the performance in each excerpt. Within each trial/competition of three featured performers, the SD in evaluations of quality of performance as assessed through sound was not significantly different from the SD in evaluations of quality of performance as assessed through visuals (all Ps $>$ 0.05). The lack of significance was not due to floor or ceiling effects.

At-chance rates with video-plus-sound recordings. In contrast to how participants performed at rates significantly below chance with sound-only recordings, and how participants performed at rates significantly above chance with video-only recordings, both novices (experiment 3) and experts (experiment 5) were similarly at chance in identifying the actual winners when they were randomly assigned to the video-plus-sound condition.

Although it may have been expected that more information in the form of video-plus-sound recordings would have led to higher rates of winner identification, such recordings instead led to an identification rate between those of the two conditions that included only one modality. Given that participants report the belief that acoustic information is more important in their judgment of music performance, it was likely that, when sound was made available in the video-plus-sound condition, participants relied primarily on sound. Indeed, participants did not perform at rates significantly different from those of participants who received sound-only stimuli.

This nuanced interpretation of the findings suggests that not only does visual information impact our perception of music, but also that it can dominate our perception such that it interferes with auditory information. This explanation may help account for why it appears that experts receiving video-plus-sound excerpts in the current experiments have decreased rates of identifying the winners.

This area needs further investigation in future work, but preliminary data appear to support the interpretation offered above. Data from ongoing research indicate that, when placed under cognitive load, participants in the video-plus-sound condition—for the first time—identify the winners at rates significantly above chance. In natural evaluation settings such as music competitions, perhaps a visual primacy emerges as our attention is exhausted and we become less able to focus on using sound as the primary information. The ongoing research suggests that we may revert to a dependence on visual information when overwhelmed with information.

On the other hand, our conscious recognition of the importance of sound may guide our attention toward sound when we face more manageable sets of information. In the video-plus-sound condition in experiments 3 and 5, it may be that the limited evaluation setting allowed people to focus on sound. However, this condition ironically led to decreased rates of identifying the winner, perhaps because the availability of sound led people away from the original (visually based) decisions.

Experiment 5. These findings are in line with other work on rating behavior that showed a strong relative similarity in response behavior between experts and novices (1). Building on the earlier work, the current manuscript offers a more detailed scope of inquiry by assessing whether expertise affects the likelihood that judgments are impacted by each modality.

Experiments 6 and 7. With silent videos, participants cited the following main visual cues that led them to their choices: movement, energy, passion, coordination, intensity, effort, skill, technique, difficulty of repertoire, posture, presence, involvement, rhythm, style, precision, confidence, enjoyment, control, and consistency.

Future research will expand upon the current work and will include coders blind to the outcomes who evaluate the performances and content analyses of evaluations made of the winning versus nonwinning performances. Ongoing research also targets visual attention and processing, using eye-tracking and facial expression recognition technology to delve further into the mechanisms underlying the dominance of visual cues in judgment.

SI Results: Supplementary Experiments

Experiments S1–S3. These studies supplement the main experiments and use youth competitions that allowed for greater ease of choice. Originally the first exploratory tests of the impact of visual and auditory information, the findings motivated more rigorous examinations with multiple trials and item analyses for robustness of effect.

Experiment S1 offered a test of the speed of judgment in the domain of music, examining whether costly and time-intensive screenings are necessary. Participants with little to no experience with classical music were given video excerpts from two live music competitions featuring precollege musicians. In the first trial, 53.6% of participants identified the actual winner chosen out of six candidates by a panel of expert judges, a proportion significantly higher than predicted by chance, $\chi^2(1, n = 56) = 10.32$, $P = 0.001$. Similarly, in the second competition, 66.1% of participants identified the winner out of eight candidates at rates better than chance, $\chi^2(1, n = 56) = 146.94$, $P < 0.001$. Novices were able to approximate the original expert judgment with just 6 s of the performances. These findings suggest that lengthy professional investment in screening may not be warranted, given the quickness with which even novices approximated expert decisions made after much longer performances.

Experiments S2 and S3 used the stimuli from experiment S1 to provide tests of how visual cues affect judgment in the domain of music. Experiment S2 provided video recordings without sound. Here, 63.9% of novice participants identified the winner in the first competition, a proportion significantly higher than chance, $\chi^2(1, n = 36) = 15.13$, $P < 0.001$. Similarly, 44.4% of participants were able to identify the winner in the second competition, again performing significantly better than chance, $\chi^2(1, n = 36) = 33.59$, $P < 0.001$. Novices were able to approximate the original expert judgments of music competitions when presented with 6-s excerpts of performances that had been stripped of their sound.

Experiment S3 offered a direct comparison of judgment between two types of stimuli: (*i*) video only without sound and (*ii*) sound only without video. In the first competition, novice participants performed significantly above chance with both sound-only (68.7%) and video-only (69.7%) recordings, $\chi^2(1, n = 99) = 55.68$, $P < 0.001$; $\chi^2(1, n = 99) = 58.91$, $P < 0.001$, respectively. Similar patterns emerged for the second competition with both sound-only (53.0%) and video-only (56.0%) recordings, where participants were again significantly above chance, $\chi^2(1, n = 100) = 149.97$, $P < 0.001$; $\chi^2(1, n = 100) = 173.00$, $P < 0.001$.

In the judgments of youth competitions, differences in performance quality were easily perceptible through either modality. Differences in the relative degree to which judgments were influenced by visual versus auditory cues may be less likely to surface, as the choice of the winner was far more obvious. Assuming that there is some correlation between visual and acoustic information, and also some nonoverlapping information between the two modalities, high variance trials would allow multiple pathways for participants to arrive at the same conclusion regarding which musician was clearly performing at a higher level than the rest. In support of this hypothesis, the rates of identifying the actual winner were relatively high in either condition.

Experiments S4 and S5. Experiment S4 and S5 examined potential effects of static visual information. Although the outline videos presented coarse visual information, demographic cues such as race and sex may still be available. However, when participants were presented with still photographs of the contestants, they were not able to select the actual winners at rates significantly above chance (36.8%), $\chi^2(1, n = 48) = 2.75$, $P =$ n.s. By extrapolation, visible but static demographic cues did not significantly impact professional judgment in the actual competitions.

Experiment S5 examined the effects of physical attractiveness on the expert judgment of quality of musicians. In this experiment, novice and expert participants were asked to identify the most physically attractive contestant upon viewing silent videos of their performances. Their choices were at chance (32.1%) in comparison with the actual winners of the competitions, $\chi^2(1, n = 38) = 0.26$, $P =$ n.s. If the original expert judges had been influenced by physical attractiveness, these evaluations would have been significantly above chance.

SI Materials and Methods

Experiment 1. When no tax was placed on the video-plus-sound option, the vast majority of participants would choose to receive this option to maximize their likelihood of selecting the actual winners of the competitions. As the original conditions of the live competitions included both visuals and sound, the video-plus-sound recordings offered both more information and a better approximation of the original conditions under which the decisions were made. As obtaining more information was costless under these conditions, participants were far more likely to choose the video-plus-sound recordings.

The tax on the video-plus-sound recordings, the option with most information, thus offers a strong test of beliefs about the judgment of performance. Given that most participants wish to maximize their study earnings, the tax forced them to consider how much additional information was worth, in allowing them to increase their likelihood of choosing the actual winners.

Most participants did not appear to believe that having visual information was worth the small tax on potential bonus earnings. The data suggest that significantly more participants believe that sound is more relevant than visuals to the judgment of music; the incentives built into this experiment suggest that these behavioral choices are truer indicators of beliefs about the importance of sound in this domain, and not simply self-reported beliefs that may be more subject to social norms, impression management, etc.

Experiment 2. Stimuli were excerpted from publicly available recordings from these international competitions: the Van Cliburn International Piano Competition, the International Tchaikovsky Competition, the Queen Elisabeth International Music Competition of Belgium, the International Franz Liszt Piano Competition, the Cleveland International Piano Competition, the Hannover International Violin Competition, and the San Marino International Piano Competition.

Excerpts. The length of 6 s was chosen based on the previous literature using thin slices (2). The excerpts were selected based on several criteria: (*i*) if the finalists performed the same composition during the last round of competition, identical excerpts were selected from each contestant; (*ii*) if the finalists performed different compositions, excerpts were selected such that the excerpts showcased approximately equal technical difficulty and musical intensity.

The excerpts selected and used controlled for the position from which the footage was recorded. Within each competition set, the contestants were captured on film from comparable distances and positions. The excerpts were selected such that the winners were not particularly favored or featured through close-ups that would have revealed facial expressions at rates higher than for the nonwinners.

All excerpts were pretested on a separate sample of 29 professional musicians with an average of 16.48 y of formal training; repeated-measures ANOVAs with post hoc pairwise comparisons suggested that, among each competition trial, there were no singular excerpts that were significantly distinguishable from the others on the dimensions listed above ($Ps > 0.05$). Throughout all experiments, stimuli were presented in random order: (*i*) randomization among sets of competitions, and (*ii*) randomization in performer excerpts within competitions. Participants were not able to match the silent video-only versions of the recordings with the sound-only versions of the same recordings.

The item analysis conducted involves a trials testing model to assess whether effects are driven primarily by certain specific trials. This technique is often used in test construction to examine whether test items are comparable in characteristics such as content and form. To demonstrate in more detail, in the initial tests, the analyses by participants average across trials and compare the participant average across all trials in condition A against the participant average across all trials in condition B. The item analysis supplements the previous analyses, averaging across participants and comparing the item average across all participants in condition A against the item average across all participants in condition B.

Explicit beliefs. The order varied regarding at which point explicit beliefs were elicited about the importance of acoustic versus visual information. In experiment 1, along with several dozen surveys not included in the current manuscript due to space constraints, these choices were made before the receipt of any recordings.

In other experiments, these questions were included along with basic demographic items at the end of the studies. If the explicit beliefs reported had been (temporarily) affected by the process of receiving the excerpts, these responses would have differed depending on the condition to which participants had been randomly assigned. Supplementary analyses using χ^2 tests suggest that this alternative explanation did not account for the findings, $\chi^2(4) = 0.899$, $P = 0.925$. There is no statistically significant association between condition and which modality was chosen as most important in judgment. Regardless of condition, participants were much more likely to have selected acoustic information as more important in the judgment of music performance.

Experiment 3. The apparent dissociations between what is reported as valued and what is actually used is not due to cynical attributions about the original judges' objectivity, motivation, or abilities. In a supplemental experiment, using the same recordings and between-subjects design, 69 participants were asked, "Who do you think should win the competitions?" instead of "Who won the competitions?" The results replicated the patterns from the previous studies, and participants' own choices approached the actual outcomes solely in the silent video-only condition, with 44.2% selecting the actual winners, significantly above chance, $t(21) = 2.19$, $P = 0.040$. They were below chance (23.5%) with sound-only recordings, $t(24) = -3.54$, $P = 0.002$, and at chance (29.3%) with recordings with both video and sound.

Experiment 4. To control for expert recognition of specific musicians, data from participants who recognized any musicians through the video-only recordings were discarded.

Experiment 5. To control for recognition of musicians, data from participants who recognized any musicians through the video-only and video-plus-sound recordings were discarded. A test of proportions found that in terms of sex and ethnic breakdowns, the ratios of performers selected as winners to all finalist performers was not significantly different from what would be expected from the population of finalists. Furthermore, there were no significant differences due to the age, sex, or ethnicity of the participants. No main effect of age, sex, or ethnicity on ability to identify the actual winners emerged, and there were no significant interactions between the demographic variables and the assigned conditions. There were also no significant dyadic effects, such as from homophily between participant and performer. For example, female participants were not more likely to prefer female musicians over male ones, and Asian participants were not more likely to choose Asian musicians as the winners.

Experiment 6. The first experiments on the phenomenon in which visual information is privileged above auditory information in the judgment of music motivated an investigation of mechanisms. As gestures, movement, and expression were cited in free-response data provided by participants when they described the type of information they relied on when making their decisions, experiment 6 explored whether impressions made based on gestures alone would approximate the decisions made by the original judges under live-round competition conditions.

The literature on music communication models (3) provides rich areas of discussion, highlighting the ways in which meaningful communication in performance often includes highly expressive movements. This work has investigated how body movements and facial actions contribute to the production of expressive performance, are used for purposes of expressive effects, and relate to and communicate with coperformers and audience members. Each aspect of this literature holds important relevance for the continued exploration of how and why visual information appears to have so significantly influenced professional judgment.

SI Materials and Methods: Supplementary Experiments

Experiment S1. Fifty-six participants ($M_{\text{age}} = 22.09$, SD = 2.16; 57.1% male) with little to no experience in classical music volunteered.* Participants received excerpts with both sound and video from two competitions held at a mid-Atlantic conservatory. They were presented with the first 6 s of the performances and were asked to identify the winners.

All contestants had performed for 10–15 min during the original live performance competitions, which included both sound and video for the actual expert judging panel. All contestants performed on the piano in a public venue. The original judging panel consisted of conservatory faculty and other internationally noted musicians. Within each competition, the excerpts were presented in random order.

In each trial or set of competition stimuli, external constraints due to competition rules and regulations provided some level of control. For example, in the youth competitions, all contestants had similar levels of formal expertise, being required to be no older than the age of 12. Their level of performance experience would have

*Participants were recruited from a community sample in the northeastern United States and were paid $5 for their participation.

also been similar, as the competitions were based in a conservatory precollege program that required entrance auditions.

Two stimuli sets were used, and the main difference was the number of pianists who chose to participate in each competition. For these initial exploratory tests, because the variance in level of performance was expected to be higher given that youth musicians were featured, all available target performers (controlling for instrument) were included in the stimuli sets.

Experiment S2. Thirty-six participants (45.2% male) with little to no experience in classical music volunteered.* Participants received silent video versions of the experiment S1 stimuli. They were then asked to identify which performers won the competitions. Within each competition, the excerpts were presented in random order.

Experiment S3. One hundred participants ($M_{\text{age}} = 23.59$, SD = 2.82; 42.7% male[†]) with little to no experience in classical music volunteered.[‡] Through a within-subjects design, each participant received both the video-only set and the sound-only set of the same performances. Using the experiment S1 and S2 stimuli, the two versions of recordings featured the same 6 s of performance for each musician, ensuring identical performance content and quality.

For example, a participant might see silent video-only recordings of eight contestants in one competition and be asked to identify the winner; later, the same participant would hear sound-only recordings of the same eight contestants and be asked to identify the winner. Participants were not able to match the silent video-only versions of the recordings with the sound-only versions of the same recordings. Any potential influence of one version on a subsequent version should not have significantly impacted the overall effects, given that the order of presentation of the two different conditions was counterbalanced. Within each competition, excerpts were presented in random order.

Experiment S4. Forty-eight participants ($M_{\text{age}} = 26.58$, SD = 9.37; 41.7% male) volunteered.[‡] Participants received still photographs of the musicians from the experiment 2–5 stimuli and were then asked to identify the winners of each competition.

The use of still photographs in this study for assessments of physical attractiveness is similar to that used in recent work (4). Although the earlier research suggests that physical attractiveness may mediate the relationship between pop music performers and participant aesthetic responses, physical attractiveness is less explicitly accepted as legitimately contributing to judgments about classical music. For example, there may be more variance considered acceptable within the pop genre in terms of performance attire whereas strong norms remain regarding adherence to traditional concert dress in the classical genre. In addition, the earlier work investigated adolescent subjects whereas the current research focuses on the expert judgments of professional musicians, whose training and experience may render them less subject to the influence of physical attractiveness.

Experiment S5. Thirty-eight participants ($M_{\text{age}} = 21.92$, SD = 4.21; 41.7% male[†]) volunteered.[‡] Participants received the video-only versions of the experiment 2–5 stimuli and were asked to identify the most physically attractive contestants.

[†]Participants who did not report their sex were not included in the calculation.

[‡]Participants were recruited from a community sample in the northeastern United States and were paid $20 for their participation in an hour-long set of unrelated studies that included the current experiment.

1. Thompson S (2006) Audience responses to a live orchestral concert. *Music Sci* 10(2): 215–244.
2. Ambady N, Rosenthal R (1993) Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J Pers Soc Psychol* 64(3): 431–441.
3. Davidson JW, Malloch S (2009) *Communicative Musicality: Exploring the Basis of Human Companionship*, eds Malloch S, Trevarthen C (Oxford Univ Press, New York), pp 565–583.
4. North AC, Hargreaves DJ (1997) The effect of physical attractiveness on responses to pop music performers and their music. *Empir Stud Arts* 15(1):75–89.

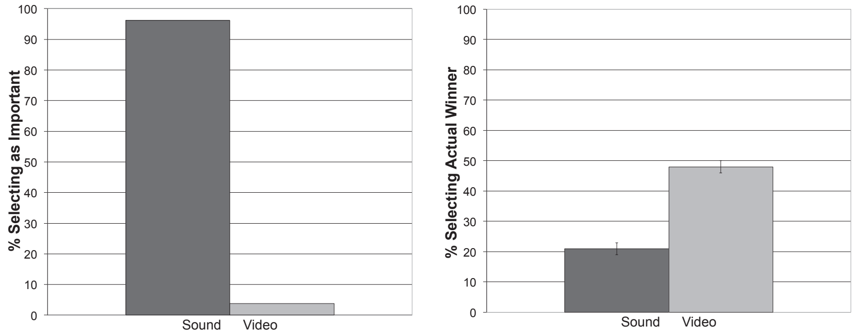

Fig. S1. A comparison of the reported importance of sound versus visuals for evaluation (*Left*), with the % identifying actual outcomes when given sound-only versus video-only stimuli (*Right*), in experiment 4 ($n = 35$). Using a within-subjects design, this study tested the impact of visual information on professional musicians.

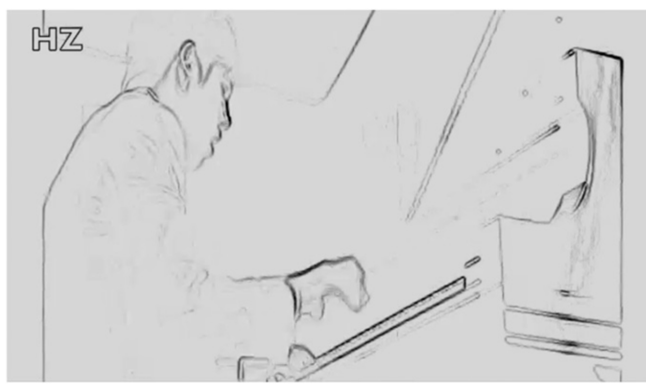

Fig. S2. Sample outline figure used in experiment 6, isolating visual information to basic motion alone. The outlines are the detected regions/silhouettes of movement. After receiving silent performance excerpts of the musicians as rendered in the above example, participants were asked to identify the winners of each competition.

Table S1. Summary of experiments

Exp.	N	Participant type	Stimulus type	Versus at chance	Against other conditions
1	106	All	Choice of recording	V: 14.2% A: 58.5% V/A: 27.4% $\chi^2(2, n = 106) = 32.96$, $P < 0.001$, $\omega = 0.558$	V vs. A: $\chi^2(1, n = 77) = 28.89$, $P < 0.001$, $\omega = 0.613$ V/A vs. V: $\chi^2(1, n = 44) = 4.46$, $P = 0.035$, $\omega = 0.318$
2	106	Novice	Professional competition	V: 52.5%, $t(105) = 10.90$, $P < 0.001$ A: 25.5%, $t(105) = -5.23$, $P < 0.001$	V vs. A: $t(105) = 12.07$, $P < 0.001$; Cohen's $d = 1.66$, 95% CI [1.63, 1.69]
3	185	Novice	Professional competition	V: 46.4%, $t(49) = 4.04$, $P < 0.001$ A: 28.8%, $t(66) = -2.09$, $P = 0.040$ V/A: 35.4%, $t(67) = 0.94$, $P =$ n.s.	V vs. A: $t(115) = 4.68$, $P < 0.001$; Cohen's $d = 0.88$, 95% CI [0.82, 0.93] V vs. V/A: $t(116) = 2.93$, $P = 0.004$ A vs. V/A: $t(133) = -2.14$, $P = 0.034$ [adjusted $\alpha = 0.017$]
4	35	Expert	Professional competition	V: 46.6%, $t(34) = 4.05$, $P < 0.001$ A: 20.5%, $t(34) = -6.11$, $P < 0.001$	V vs. A: $t(34) = 5.89$, $P < 0.001$; Cohen's $d = 1.42$, 95% CI [1.36, 1.46]
5	103	Expert	Professional competition	V: 47.0%, $t(32) = 3.40$, $P = 0.002$ A: 25.7%, $t(29) = -3.34$, $P = 0.002$ V/A: 29.5%, $t(39) = -1.43$, $P =$ n.s.	V vs. A: $t(61) = 4.48$, $P < 0.001$; Cohen's $d = 1.09$, 95% CI [1.01, 1.14] V vs. V/A: $t(71) = 3.72$, $P < 0.001$ A vs. V/A: $t(68) = -1.00$, $P =$ n.s. [adjusted $\alpha = 0.017$]
6	89	All	Professional competition	48.8%, $t(88) = 6.49$, $P < 0.001$	N/A
7	262	All	Professional competition	V_{Conf}: 37.4%, A_{Conf}: 39.5% V_{Crea}: 44.6%, A_{Crea}: 26.1% V_{Inv}: 53.1%, A_{Inv}: 34.2% V_{Mot}: 52.8%, A_{Mot}: 35.6% V_{Pass}: 59.6%, A_{Pass}: 38.7% V_{Uni}: 43.6%, A_{Uni}: 26.3%	$t(193) = -0.68$, $P =$ n.s. $t(260) = 9.00$, $P < 0.001$ $t(260) = 9.60$, $P < 0.001$ $t(260) = 7.91$, $P < 0.001$ $t(196) = 7.01$, $P < 0.001$ $t(192) = 6.22$, $P < 0.001$ [adjusted $\alpha = 0.010$]
S1	56	Novice	Youth competition	V/A^1: 53.6%, $\chi^2(1, n = 56) = 10.32$, $P = 0.001$, $\omega = 0.429$ V/A^2: 66.1%, $\chi^2(1, n = 56) = 146.94$, $P < 0.001$, $\omega = 1.620$	N/A
S2	36	Novice	Youth competition	V^1: 63.9%, $\chi^2(1, n = 36) = 15.13$, $P < 0.001$, $\omega = 0.648$ V^2: 44.4%, $\chi^2(1, n = 36) = 33.59$, $P < 0.001$, $\omega = 0.966$	N/A
S3	100	Novice	Youth competition	V^1: 69.7%, $\chi^2(1, n = 99) = 58.91$, $P < 0.001$, $\omega = 0.771$ A^1: 68.7%, $\chi^2(1, n = 99) = 55.68$, $P < 0.001$, $\omega = 0.750$ V^2: 56.0%, $\chi^2(1, n = 100) = 173.00$, $P < 0.001$, $\omega = 1.315$ A^2: 53.0%, $\chi^2(1, n = 100) = 149.97$, $P < 0.001$, $\omega = 1.225$	Individual trials
S4	48	All	Professional competition	36.8%, $\chi^2(1, n = 48) = 2.75$, $P =$ n.s.	N/A
S5	38	All	Professional competition	32.1%, $\chi^2(1, n = 38) = 0.26$, $P =$ n.s.	N/A

Several SI experiments included χ^2-based significance testing because analyses were conducted on individual trials (two trials or competitions were tested in the earlier studies). As the initial exploratory studies, analyses were conducted on each separate trial, resulting in nominal data of whether or not the actual winner was selected by participants. The main experiments included t tests because analyses were conducted on average identification rates (%) across 10 trials per participant. A, sound only; V, video only; V/A, video plus sound; N/A, not applicable.