Sight over sound in the judgment of music performance

Chia-Jung Tsay

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.

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.
(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 (SI Text). 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).
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; thus, 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, $t(34) = 5.89, P < 0.001$; Cohen’s $d = 1.01$ (Fig. S1). An item analysis indicates that this effect held across all 10 competitions, $t(9) = 3.74, P = 0.005$.

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, compared with those who heard sound-only stimuli, $t(32) = 3.40, P = 0.002$. Experts were significantly better with video-only stimuli than with sound-only stimuli, $t(61) = 4.48, P < 0.001$; Cohen’s $d = 1.20$. An item analysis indicates that these effects were robust across all 10 competitions, $t(9) = -2.36, P = 0.04$.

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

![Graph](pnas.1703956111.png)

**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 perfor-
mances. They were then asked to identify the most confident, cre-
ative, 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. Creativ-
ity, involvement, motivation, passion, and uniqueness were signif-
ically 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 judgments (62). The 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 con-
tribute to the professional judgment of music. Motion, motivation,
creativity, and passion are perceived as hallmarks of great per-
formance (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’ move-
ments may contribute substantially toward inferences about
the quality of performance. Our movements facilitate aspects of cog-
nitive 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 in-
tuition that musicians’ motions and emotions represent excep-
tional 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 automati-
cally 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-based 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 per-
formance evaluation in a multiplicative cross-modal model of per-
ception (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 cur-
rent work focuses on choices made during competitive settings,
more information would not necessarily lead to better approx-
imations 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 in-
formation is available or not. This finding suggests that visual
cues are indeed persuasive and sway judges away from recog-
nizing the best performance that they themselves have, by con-
sensus, 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 imple-
menting incentives and bonuses for participant performance in
identifying the actual winners. However, both experts and nov-
ices 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 know-
ing 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 judg-
ment 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 perfor-
mance, whether it concerns hiring employees, interviewing physi-
cians, 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 deci-
sions. 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.
Experiment 1. One hundred six participants (M_\text{age} = 20.73, SD = 2.46; 49.5% male\textsuperscript{*}) volunteered.\textsuperscript{7} 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 $10 if their selections matched the actual competition outcomes. They had performed sound or video recordings; or, if they chose to record, with both video and sound, $2 would be deducted from any bonuses won.

Experiment 2. One hundred six participants (M_\text{age} = 22.26, SD = 1.79; 41.1% male\textsuperscript{*}) with little to no experience in classical music volunteered.\textsuperscript{7} 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_\text{age} = 24.18, SD = 9.64; 46.1% male\textsuperscript{*}) with little to no experience in classical music volunteered.\textsuperscript{7} 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 5. Eighty-nine participants (M_\text{age} = 27.38, SD = 10.68; 50.0% male\textsuperscript{*}) volunteered.\textsuperscript{7} Participants received silent videos from the experiment 2–5 stimuli that had been reduced to black-and-white moving outlines (Fig. 52). Participants were then asked to identify the winners of each competition.

Experiment 6. Two hundred sixty-two participants (M_\text{age} = 21.52, SD = 3.36; 52.3% male) volunteered.\textsuperscript{7} 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 of the final sets. Repeat choices were allowed.

ACKNOWLEDGMENTS. I thank M. Bazermark 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 N. Nakayama for the forum that sparked the initial studies; J. Abraham, R. Cheong, S. Dobrow, P. Filli, A. Galinsky, J. Gollan, M. Morgan, O. Obodaru, D. Parno, K. Shonk, and the University College London M Sl 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.