The Communication of Timbral Intentions Between Pianists and Listeners and Its Dependence on Auditory-Visual Conditions

Shen Li[1,2], Renee Timmers[3] and Weijun Wang[1,2]*

[1] *School of Psychology, Central China Normal University, Wuhan, China,* [2] *Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan, China,* [3] *Department of Music, The University of Sheffield, Sheffield, United Kingdom*

The perceptual experiment reported in this article explored whether the communication of five pairs of timbral intentions (bright/dark, heavy/light, round/sharp, tense/relaxed, and dry/velvety) between pianists and listeners is reliable and the extent to which performers' gestures provide visual cues that influence the perceived timbre. Three pianists played three musical excerpts with 10 different timbral intentions ($3 \times 10 = 30$ music stimuli) and 21 piano students were asked to rate perceived timbral qualities on both unipolar Likert scales and non-verbal sensory scales (shape, size, and brightness) under three modes (vision-alone, audio-alone, and audio-visual). The results revealed that nine of the timbral intentions were reliably communicated between the pianists and the listeners, except for the dark timbre. The communication of tense and relaxed timbres was improved by the visual conditions regardless of who is performing; for the rest, we found the individuality in each pianist's preference for using visual cues. The results also revealed a strong cross-modal association between timbre and shape. This study implies that the communication of piano timbre is not based on acoustic cues alone but relates to a shared understanding of sensorimotor experiences between the performers and the listeners.

Keywords: music communication, timbre perception, performer-listener chain, auditory-visual condition, cross-modal correspondence

INTRODUCTION

Compared to the communication of emotions or expressiveness in musical performance, research conducted into the communication of timbral intention between performers and listeners is rare. From the performers' perspective, they intend to produce explicit timbres in their performances, which may lead to their self-satisfaction and sense of achievement (Holmes, 2011), relate to their holistic perception of expressive elements and music structure (Li et al., 2020), or be relevant to emotional expression (Juslin, 2000). In addition, piano teachers and students usually work on explicit timbral intentions in piano lessons using metaphors, gestures and modeling etc. to improve their communication of timbre-related performance goals (Li and Timmers, 2021). However, whether pianists can communicate timbral intentions reliably to the listeners under different auditory-visual presentation conditions remains unknown. An interview study (Li et al., 2020) on the conceptualization of piano timbre revealed pianists' extensive utilization of timbral intentions in piano performance; the findings also suggested that a pianist's concept of timbre is enriched by embodied experience, such as bodily preparations, indicating the relevance of visual

cues. As a follow-up study, a perceptual experiment was conducted to examine the accuracy of communication of timbral intentions to listeners and its dependence on the visual and aural components of musical performance. It aims to explore the core question of how piano timbre is communicated, by using several sub-questions: Do pianists communicate timbral intentions to the audience and is the communication reliable? And what is the relevance of the auditory and/or visual components of a musical performance in timbre communication?

THE RESEARCH BACKGROUND OF PIANO TIMBRE

To better understand the research background and research purpose, it is necessary to clarify the meaning and scope of timbre specified in this study, which we distinguished as "micro-perspective" and "macro-perspective." In general, timbre studies have usually adopted a macro-perspective, which considers the differences in timbre produced from various sound objects (i.e., different sound sources). The "macro-perspective" focuses on the characterization of timbre of specific instruments or instrument groups. This could also concern differences within an instrument group (i.e., the timbre of piano A is different from that of piano B). An important aim may be to uncover the timbre space through which listeners categorize or distinguish instrument timbres (e.g., timbre intervals: McAdams, 2013). However, this study investigates piano timbre from a "micro-perspective"—which focuses on the timbral nuances produced from one instrument, from the point of view that a specific instrument can still have a variety of timbres depending on how the instrument is played (e.g., piano A has contrasting timbres when the key is either gently pressed or quickly struck).

One difficulty in the study of piano timbre lies in the interweaving of timbre and other performance parameters (e.g., intensity, articulation, and tempo) in the pianists' conceptions of piano timbre. It might be difficult for pianists to believe that, by controlling performed intensity, different ways of touching the keys have little effects on produced timbre—what they've changed are the attack noises (i.e., finger-key noise, key-keyframe noise; Goebl et al., 2014). Timbre may also be phrased as tone quality, tone color in the discourses of pianists.[1] Sometimes we found these terminologies were used interchangeably in the writings and discourses of pianists, with the combined effect of timbre and other musical features being mentioned. Ortmann (1935) referred to "tone quality" and suggested that the perception of tone quality is subjective and results from our unified reaction to three variants: pitch, intensity, and duration. Bernays and Traube (2014) put forward the notion of *composite timbre*, referring to the complexity of piano timbre interwoven with other performance parameters when considering piano timbre in a musical and polyphonic context. Their study selected five verbal descriptors (dry, bright, round, velvety, and dark) and the pianists conceptualized and performed the given music pieces in accordance with these timbral nuances. For these

reasons, this study investigates the beliefs and utilizations of piano timbre within the conventions of the pianistic community, rather than from an acoustic perspective. In other words, we will examine timbral communication in the context that pianists conceptualize their entire performance as expressing a timbral feature (i.e., mental conception: Kochevitsky, 1967) and rely on the coordination and adjustment of other musical parameters to align with the timbral intention.

MUSIC COMMUNICATION BETWEEN PERFORMER AND LISTENERS

The Role of Visual Information

Previous studies that investigated the role of visual information provided by the performer in communicating to the listeners mainly consider the communication of expressivity (Davidson, 1993, 1995; Broughton and Stevens, 2009) and emotional intentions (Dahl and Friberg, 2007). In Davidson's study (1993), a pianist was instructed to play a piece of music in three differently expressive ways (deadpan, projected, and exaggerated) and listeners were asked to rate the *expressivity* in the condition of seeing only, hearing only, or both seeing and hearing. This study found that once the pianist was playing with a certain degree of expressivity (i.e., in a projected and exaggerated manner), listeners could only differentiate between different degrees when the visual information was present and not in the audio-only presentation. Employing a similar experimental paradigm, Broughton and Stevens (2009) verified this result in the communication of expressivity in a marimba performance in which the audio-visual condition helped the listeners' differentiation of performances played with different expressive intentions compared to the audio-only condition. These studies suggested that the bodily movements and gestures of the performers provide identifiable information through which listeners can detect musical expressivity.

Performers express and communicate emotions to audiences [see a review by Juslin and Timmers (2010)] and visual information plays a role in the performer-audience communication chain (Camurri et al., 2003; Timmers et al., 2006; Dahl and Friberg, 2007) and for a duo partner (Wöllner, 2020). Dahl and Friberg (2007) examined whether the specific emotions (happy, sad, angry, and fear) expressed by the performer could be recognized by listeners using visual information only. This study confirmed that listeners (not necessarily musically trained) could easily recognize emotions such as happy and sad and were roughly accurate with anger, but failed to perceive fear. In a dancing-related study, Camurri et al. (2003) indicated that listeners can successfully detect a dancer's emotional intentions (e.g., joy, fear, grief, and anger) through movement cues only. These studies implied that even without the aural cues, visual cues are informative for listeners to recognize the emotional intentions of performers.

However, to the author's current knowledge, there is no existing research that investigates the communication of timbre in music performance and considers the visual effects of performers' gestures in this process. A possible exception may

---

[1]See Li and Timmers (2021) and Li et al. (2020) for a detailed explanation of the similarities and differences in these terminologies.

be a research project by Wapnick et al. (2004) who examined the visual perception of tone quality. The results indicated that the ratings of perceived tone quality together with the other five performance items (note accuracy, rhythmic accuracy, expressivity, adherence to style, and overall impression) were higher in the audio-visual presentation mode than the audio-only presentation. The present study aims to examine whether the communication of timbral intention between the pianist and the listeners is reliable and to what extent visual information influences judgment.

Experimental Paradigm
Changing Audio-Video Presentation Modes
The first common approach in empirical studies of visual communication in music performance is to: (a) instruct the performer to play the same piece in a different manner (expressive vs. non-expressive, or with different emotional intentions) and (b) vary the mode of audio-visual presentation of the recordings and investigate its influence on the listeners' evaluation of the performance. Relevant research has been conducted to investigate the role of vision in the evaluation of emotional engagement (Timmers et al., 2006), performance quality (Wapnick et al., 2000, 2004), and expressive intention (Davidson, 1993, 1995). Similarly, visual information from a singer's facial expression (e.g., eyebrow raised or not, eyes widened or not, lip movement) affected a listener's judgment of the emotional valence of a music piece (Thompson et al., 2005).

Employing Congruent/Incongruent Pairings
Another approach employed by researchers is to make congruent or incongruent pairings of audio-video stimuli and examine their impact on listeners' responses. This approach helps to understand the impact of auditory-visual integration on participants' processing of new music stimuli, as the "McGurk" effect demonstrated in the studies of speech communication (McGurk and MacDonald, 1976). Relevant research has been conducted to investigate the extent to which vision influences the processing of tone duration (Schutz and Lipscomb, 2007), pitch interval (Thompson et al., 2005; Thompson and Russo, 2007), and timbre (Saldaña and Rosenblum, 1993). For example, Thompson et al. (2005, Experiment 3) examined the influence of (in)congruent pairings of singers' facial expressions with singing clips on the perception of interval size and found that seeing an incongruent pairing (e.g., a small melodic interval accompanied by images of singing a larger interval) resulted in ratings of a larger interval than either of the original pairings. The perceptual experiment conducted by Schutz and Lipscomb (2007) demonstrated that the visual perception of stroke action (seeing either longer or shorter physical gestures) in a marimba performance influences listeners' perception of tone duration. In addition, Saldaña and Rosenblum (1993) examined the perception of cello timbre influenced by the visual information of stroke action (plucking/bowing) in cello performances. Their studies found that watching a video performance enabled the listeners to distinguish between plucking and bowing and that seeing a bowing movement led to higher ratings of bowing

timbre and seeing a plucking action resulted in a larger plucking timbre response.

The above studies suggest that we could either manipulate the presentation mode of one stimulus, or combine two stimuli that convey different timbres. This study will use the first approach (changing aural-visual presentation modes) to examine the visual communication of timbral intentions between pianists and listeners, due to the ease of operation and the study purpose. Playing with different timbral intentions leads to a variety of performance parameters e.g., timing and performance speed, which result in the difficulty of synchronizing audio with unmatched video in the "artificial" audio-visual stimuli (i.e., incongruent pairing). Instead, the second approach will be more effective in examining the perception of single piano tones or chords, but the focus of this research is the polyphonic musical background.

There are increasing numbers of scholars (Dahl and Friberg, 2007; Behne and Wöllner, 2011) investigating in detail which visual aspects and which specific regions of bodily movements have an impact on listeners' music perception, instead of looking performers' bodily movements as an integrated part. For example, ancillary gestures—those that are not required for producing sounds (Wanderley et al., 2005; Wanderley and Vines, 2006), have been found to have an impact on audiences' perception of musical tention and phrasing (Vines et al., 2006). Sound-producing gestures—those movements that are effectively produce sounds (e.g., the hands of a marimba player, or the lips of a singer; Jensenius et al., 2010), can influence listeners' judgement of musical notes at a perceptual level (see section Employing Congruent/Incongruent Pairings). Dahl and Friberg (2007) tested the influence of viewing condition (full, nohands, nohead, and only-head) on the detection of emotions. They found that happy emotion was better detected when seeing upper body movement while sad emotion was better recognized when the head movement information was available. Seeing the whole body led to the highest accuracy of recognized emotions. These studies suggested that we could combine two angles in the video recordings—one focused on the movement of the upper body and the head, and the other one focused on the sound-producing gestures (i.e., finger movements).

AN EMBODIED PERSPECTIVE ON THE COMMUNICATION OF TIMBRAL INTENTIONS

From an Information-Processing Model to an Embodied Model
The rationale of communicating a timbral intention from the performer to the listeners can be explained by several well-known models, for instance, the information-processing model (Kendall and Carterette, 1990) and the lens model (Juslin, 1998, 2000; Juslin and Lindström, 2010). According to Kendall and Carterette (1990), the information-processing model of music communication addresses how the composer and performer convey their expressive intentions and how the listener perceives the intended message from the composer/performer *via* message

recoding. The composer's encoding process goes from ideation to notation; the performer then recodes the notation into acoustic signals and, finally, these acoustic signals are recognized by listeners and result in ideation. In this context, timbral intentions work as a message that the performer intends to communicate to their audience and the encoding and recoding of this timbral message relies on a successful recognition of acoustic cues. For acoustic cues, Juslin and Lindström (2010) adapted the lens model from seminal work Brunswik's (1956) and applied it to the musical context. The modified lens model explains the utilization of cues (i.e., tempo, loudness, articulation etc.) in the communicative process of specific emotions between the performer and the listeners, which suggests that a successful communicative process happens when the performer's cue utilization matches that of the listeners. Both models have acknowledged the feasibility of communicating implicit messages (i.e., emotions) within the musical communication process; however, these models seem to regard musical communication purely as a sonic art, as can be seen in Juslin and Lindström's model of acoustic cues as a lens of musical expression, as well as Kendall and Carterette's model which regards the acoustic signal as transmitting the composer/performer's thoughts to the listener's ideation.

The theory of embodied cognition has prompted the understanding of listeners' musical experience, as revealed by the following research. For example, Davidson and Correia (2002) highlighted the role of bodily movements in shaping the listeners' perception of musical expressiveness. More importantly, the common knowledge of bodily experience i.e., balance, scale, force, and cycles (Johnson, 1987) forms the common ground on which performers and audiences can communicate and appreciate the musical work (Davidson and Correia, 2002). Leman (2008) extended the previous research and proposed a model of embodied music communication based on the decoding and encoding of motor control. The performer realizes the musical goal/idea through corporeal articulations, while the instrument is the mediation which may form a closed loop with the performer with haptic, sonic and visual feedback whilst also transmitting the sonic and visual energy to the listener. Listeners can make sense of the communicative process through a mirror process i.e., a corporeal and cerebral understanding of the intended actions. Leman's (2008) model suggests multimodal sensing of music communication and extends the understanding of musical communication from the decoding/encoding of sonic forms to include those of motor control.

It is worth noting that listeners' embodied listening experience does not merely incorporate the visual perception of musical performance, it could involve more covert imitation of sound-producing actions. Godøy (2006) related the notion of gestural-sonorous object to musical sounds, referring to the phenomenon that humans would mentally trace sound as a continuous process (i.e., the onsets, contours, textures and envelops involving hands, fingers, arms etc.) in the perception or imagination of musical sounds. Cox (2016) proposed the mimetic hypothesis and paid attention to the covert mimetic behavior (MMI, mimetic motor imagery) besides the overt imitation of performance actions (MMA, mimetic motor action). He suggested that, MMI plays

important roles in the comprehension of musical sounds, for instance the instrumental timbre—despite the limitations of human voice to overtly imitate other instrumental timbres, "what matters is the attempt to emulate the sound, to feel something of what it would be like to make such sounds, and to thereby feel something of what it would be like to be an entity capable of making such sounds" (p. 32).

Sensorimotor Perceptions in Timbral Communication

The embodied model of music communication implies that timbral communication may be associated with sensorimotor perceptions, as indicated by several studies (De Poli et al., 1998, 2017; Baraldi et al., 2006). For example, De Poli et al. (2017) claimed that sensorimotor expressivity is the part of musical expressivity that is not covered by musical and emotional expressivity and that it reflects certain cross-modal correspondence (CMC) features. They further developed an FEI (friction, elasticity, and inertia) metaphor that characterizes the sensorimotor expressivity in a kinetic-energy two-dimensional space (De Poli et al., 2009) and demonstrates that participants can describe perceived expressivity *via* FEI metaphors (e.g., hard-soft, light-heavy categories).

Timbre communication may also involve CMC features (Wallmark and Kendall, 2018; Wallmark, 2019a,b). For example, timbre metaphors such as bright/dark, rough/smooth reflected cross-modal correspondence with vision and touch (Wallmark, 2019b). Timbre metaphors in CMC categories encompass "an embodied conceptual transfer process by which an auditory target domain (timbre) is understood in reference to a non-auditory source domain (vision, touch, taste, and smell)" (2019a, p. 594). Therefore, we predict that the listeners in our study will describe perceived timbral qualities in piano performances by referencing multimodal sensations.

Communicating Timbre: Emotion and Musical Expressivity

In the field of music emotion studies, timbre is closely related to the perceived emotions of listeners (Gabrielsson and Juslin, 1996; Balkwill and Thompson, 1999; Hailstone et al., 2009) as well as the experienced emotions of performers during music performance (Holmes, 2011; Van Zijl and Sloboda, 2011). For example, the selection of instrumental timbre effectively influences listeners' judgment of emotion conveyed by the music (Balkwill and Thompson, 1999). With a systematic control on other performance parameters (loudness, tempo, and melody), there is still a robust effect induced by timbral feature (e.g., spectral content, attack) on the ratings of listeners' perceived emotion (Hailstone et al., 2009). Juslin and Timmers (2010) suggested that: a bright timbre is used for expressing happiness; a soft timbre is suitable for expressing tenderness; a dull timbre is good for expressing sadness; a sharp timbre is appropriate for expressing anger. From a performer's perspective, producing timbral nuances is not only about playing techniques, but is inseparable from musicians' own emotions experienced during music practice/performance. Holmes (2011) suggested

that the employment of specific timbre is the key element of musicians' internal motivation to make a successful performance during which process their felt emotions are mainly feelings of satisfaction and elation. Van Zijl and Sloboda (2011) found that focusing on 'sound/tone color' is one of the key elements of the inner technique toward an expressive music interpretation, and the process of focusing helped them to bring their own emotions in line with the musical emotions.

The above research seems to suggest that the expression of timbral nuances, driven by affective intention of the performer, appears to be an accessory of emotional communication in musical performance. However, this study would rather consider timbral communication not under the umbrella of emotional communication, but within the framework of musical expressivity (as shown in **Figure 1**). Our previous research (Li and Timmers, 2021) found that playing a short musical phrase or an entire section with specific timbral intention without indicating any relevance to affective intention is naturalistic and feasible. The study results revealed that the piano teachers and students usually worked on an entire musical phrase with a consistent timbral intention (e.g., "lute-like timbre," "timbre of princess and earl") and employed multimodal communication strategies (e.g., modeling, verbal explanation, physical touch, gesture, etc.) to achieve these timbre-related goals. According to Bernays and Traube (2014), verbal instruction of basic emotions and timbral intentions are not comparable, and the vocabularies of describing piano timbre are consensual and meaningful to pianists (Faure, 2000; Bellemare and Traube, 2005). The study of timbre semantics (Wallmark, 2019a) suggests that the description of timbre impression can include many possibilities beside emotion, including acoustics, cross-modal correspondence (CMC), mimesis, action, etc.

Timbre, as an important musical message which conveys musical expressivity between performers and listeners, has received scholarly attention (Juslin and Laukka, 2003; Barthet et al., 2010). Barthet et al. (2010) found that changes in timbral features across expressive levels in a clarinet performance did not happen at every note but were specific to some notes or groups of notes or specific musical passages. When cellists were asked to make either expressive performance or physically constrained performance (Rozé et al., 2017), the timbre was modified in the constrained postural condition. The above studies imply that the communication of timbral intention in piano performance might involve not only affective purposes but working as an important musical message when conveying musical expressivity from performers to listeners. As mentioned before, sensorimotor expressivity is the core component of the embodied communication of musical expressivity.

The Current Research

This study aims to examine the communication of piano timbre between pianists and listeners and the influence of audio-visual presentation modes on the communication outcome. The focus on piano timbre originates from the touch-tone debate in piano performance, corresponds the trend of increasing attention on semantic studies of piano timbre (Bellemare and Traube, 2005;

FIGURE 1 | Logic relationship between timbre, emotion, and musical expressivity.
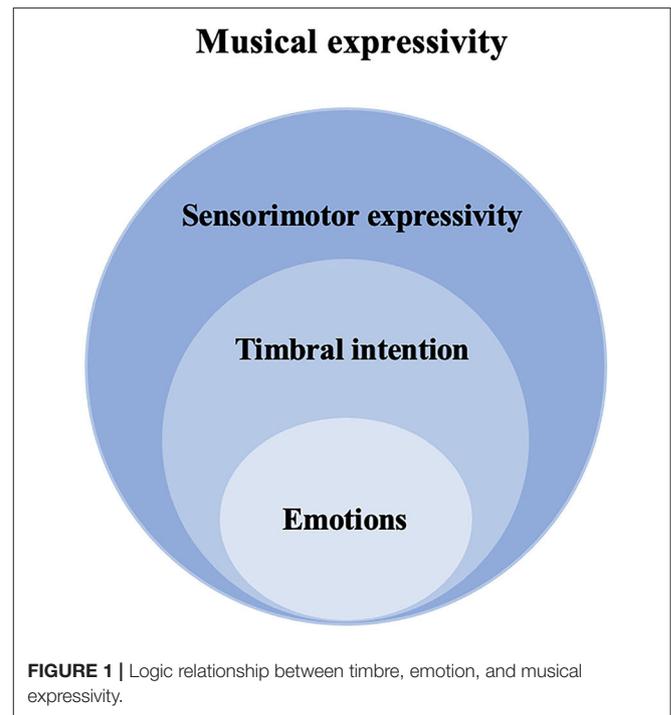
Bernays, 2013; Kojucharov and Rodà, 2015), and sheds light on the understanding of the relationship between timbre and musical expressivity (Barthet et al., 2010; Ystad et al., 2019). This research expects that, like expressivity and emotions, pianists can express different timbral intentions in their performances and listeners can detect those intentions from either the variations in sound or the visual cues in performative gestures (H1). We also expect that the listeners' ratings of perceived timbre in response to performances with contrasting timbres will be significantly different (H2). Concerning the visual influence, we expect that audio-visual presentation modes of the music performances will influence the communication of timbral intentions (H3).

Hypothesis 1 (H1): the communication of timbral intentions is reliable and successful in the performer-listener communication chain.

Hypothesis 2 (H2): the main effect of heard timbre on timbre ratings: The listeners can successfully differentiate between performances with contrasting timbral intentions.

Hypothesis 3 (H3): the interaction effect of AV condition and heard timbre: the differentiation between contrasting timbral intentions will be influenced by the auditory-visual presentation modes.

We also developed 10 timbre descriptors (see **Table 1** below) as the perceptual scales for participants in the current perceptual experiment. Bright, dark, round, velvety, and dry were selected from the study of Bernays and Traube (2014),[2] whilst the

---

[2] The validity of using the five timbre descriptors as performance instructions has been tested in their study with sound analysis and measurements of performance strategies (e.g. attack depth, speed, decay, etc.) by five pianists. Their study

TABLE 1 | Ten timbre descriptors used in the perceptual experiment.

	Timbre descriptors	
Pair 1	Bright*	Dark*
Pair 2	Round*	Sharp
Pair 3	Light	Heavy
Pair 4	Relaxed	Tensed
Pair 5	Velvety*	Dry*

*refers to the timbre descriptors borrowed from Bernays and Traube's (2014) study.

Pair A: size   Pair B: shape   Pair C: brightness
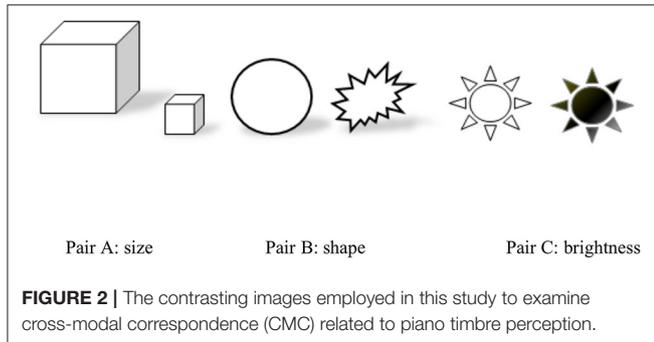
FIGURE 2 | The contrasting images employed in this study to examine cross-modal correspondence (CMC) related to piano timbre perception.

others are either their antonyms (sharp) or new descriptors in pairs (heavy/light, relaxed/tense). These 10 timbre metaphors aim to examine the multidimensionality of timbre perception. The heavy and light descriptors represent a kinetic experience associated with piano timbre perception; relaxed and tense relate to the muscular sensations when responding to piano timbre. It is noteworthy that the contrast between "velvety" and "dry" was not noticeable; however, they described the characteristics of an object's surface and represent different types of tactile feelings.

In addition to the 10 metaphors, we employed non-verbal perceptual scales (i.e., image choice) to directly test CMC features with musical timbre i.e., three pairs of images that show contrasting brightness, shape, and size (see **Figure 2**). Timbre-shape and timbre-brightness associations have been found: a soft timbre (e.g., piano) is strongly associated with round shapes together with blue, green, or lighter grayscales, whereas a harsh timbre (e.g., crashing cymbals) is associated with angular shapes together with red, yellow, or darker grayscales (Adeli et al., 2014). Additionally, timbre can be associated with visual textures that vary in terms of sharpness, compactness, and sensory dissonance (Giannakis, 2006). Therefore, this study proposes

Hypothesis 4: the timbre expressed by the performer will significantly influence the listeners' choice between contrasting images.

---

indicated that there was no difficulty in pianists interpreting the same musical piece with distinct timbral intentions, and subtle changes in expressing different timbres were found by sound and keyboard motion analysis.
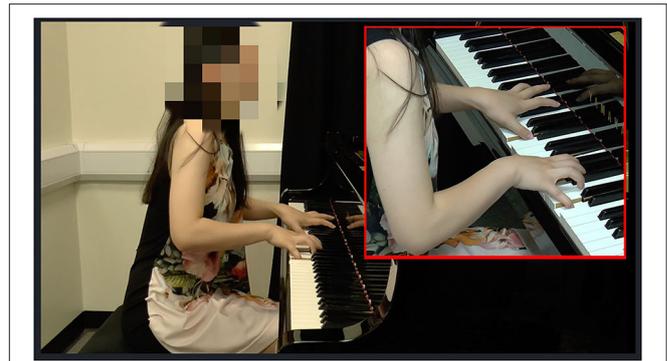
FIGURE 3 | The video presentation with two camera angles.

EXPERIMENT

Participants

Performers. Three female Chinese pianists (age 23–27) from the Department of Music at the University of Sheffield were asked to give performances to create the stimuli used in the listening experiment. They were all majoring in piano performance studies (one Ph.D. student, two Master's students) and regularly performed classical, baroque and contemporary music as part of their performance repertoire.

Listeners. Twenty-one music students (19 females and two males; Mean age = 21.89, SD = 2.03) from the music department at the University of Henan in China participated in the listening experiment. Fourteen of them were undergraduates and seven of them were postgraduates. All participants were majoring in piano performance studies. Based on large effect sizes found in the three-way interaction (Partial Eta Squared 0.15 is considered large, Cohen, 1988), we conducted a post hoc power analysis using G power 3.1 (Faul et al., 2007), that determined we had 99% power to detect large effects in a three-way interaction with 5% Type I error rate.

Stimuli

In total, there were three musical pieces x 3 AV presentations x 10 instructed timbres totaling 90 music excerpts. The three musical pieces was performed by three different pianists (piece 1 pianist 1, piece 2 pianist 2, piece 3 pianist 3), and each musical piece was played with 10 types of timbral intentions, which were audio- and video-recorded. The video recordings were captured from two angles: a global view that showed the sitting pianists from the side (i.e., the viewing of ancillary gestures), and a more local view that was focused on the hands and finger (i.e., the viewing of sound-producing gestures; see **Figure 3**).

In the preparation of experimental stimuli, the three pianists practiced and recorded all three musical pieces with 10 timbres, but only one musical piece played by each pianist was selected for inclusion in the listening experiment to avoid boredom and the effect of repetitive listening on listeners' responses (Morimoto and Timmers, 2012). Otherwise, the listeners had to listen to the same musical piece 30 times (three performers playing one musical piece repetitively using 10 different timbres). When

matching the performer and the piece, we considered the criteria of fewer mistakes and more expression in the performances of each pianist to optimize the quality of the musical material presented to listeners.

The three pianists' performances were recorded on a Yamaha grand piano (Disklavier Pro S6) in the Sound House Studio of the Department of Music at the University of Sheffield. Panasonic HC-V770k HD Camcorder and Tascam DR-05 Audio Recorder Kit were used to obtain higher recording quality.

Three pianists were given the pieces [selected from Bernays and Traube (2014), see **Appendix 1**] 1 month before the recording and each of them experienced a recording trial before the actual recording to enable them to get used to the cameras, the procedure, and the piano. The practice time also allowed them to arrive at the appropriate performance manner for each special timbre. They were told to play the piece with a consistent timbral intention and performance instructions were explained in both written (see **Appendix 3**) and oral forms. To ensure that pianists played in response to timbral intentions rather than emotional intentions, the objectives of recording their performances were clearly explained to them. The three pianists were told that their performances would be used as stimuli in a perceptual experiment and the listeners' focus would be perceived timbral quality. All three pianists have participated in the previous interview study on piano timbre, which gave them and the researchers a common understanding of the research objectives. The original musical pieces borrowed from Bernays' study do not have any expression, dynamic, articulation, phrasing, and accent markings, which gave the performers scope to vary performance parameters to achieve a desired tone quality.

Procedure

The experiment tested participants individually by presenting the recordings on a computer monitor. After the stimuli presented, 21 participants were asked to rate the presence of the 10 timbres on a 1–9 unipolar Likert scales (task 1) and to choose from three pairs of non-verbal sensory scales (task 2). We considered using unipolar scales instead of categorical responses, allowing for ambiguous and relative open responses (e.g., both bright and light might be perceived to certain extent, or none of the timbres might be strongly perceived). Each music excerpt lasted around 15 s and the participants were required to complete two tasks for each excerpt intuitively.

Participants gave their responses on a paper-based response sheet (see **Appendix 2**, two tasks in responses to each excerpt). Response sheet was written in both English and Chinese. Although there might be potential differences between Chinese and English timbre descriptors (Namba et al., 1991), the participants reported no difficulty in understanding 10 timbre descriptors which were also commonly used in Chinese context. The ordering of musical excerpts was randomized to ensure that the same timbre could not occur more than twice consecutively, and we prepared two different random orderings for counterbalancing. The monitor had a 15″ screen and was placed in front of the participants within a comfortable viewing distance (80–100 cm). The experiment contains two practice

trials and 90 experiment trials. Recordings were presented over headphones at a comfortable level. Participants were encouraged to respond intuitively after experiencing the stimuli for the first time, but they could play the recording repeatedly if they asked, or they were hesitant about the answer. The majority of them played the stimuli just once and responded quickly, but two out of 21 participants took half hour longer than others due to unsure answers and replay of stimuli. The duration of the entire experiment normally took between 45 min to an hour. In the middle of the experiment, there was a short break.

RESULTS

To examine how well the performers communicated the timbre to the participants, we calculated the percentage of correct answers in the ratings (perceived timbre) for each target timbre (heard timbre). A one-sample t-test was conducted to test whether the number of correct answers was above the chance level (10%). Timbre ratings were then entered into a three-way ANOVA with repeated measures for heard timbre (two levels: target, opposite), AV condition (three levels: AO, AV, and VO), and pianist/piece (three levels: P1, P2, and P3). To simplify the analysis and assure statistical power, the ability to communicate a particular timbre was tested per timbre pair rather than by comparing across all 10 timbre levels individually. Finally, PCA was conducted to explore which of the 10 timbres capture the same variance, with the inclusion of acoustic and visual information for further comparison.

Percent Correct Score

We calculated the percentage of correct answers for each target timbre by re-coding data as either 1 (correct) or 0 (incorrect). The percent correct scores were then further divided into an absolute percentage and a relative percentage.[3] For example, in the perception of the dark timbre, only 15.34% of listeners gave the highest ratings for darkness when exposed to an excerpt performed with *dark*[4] timbre, and the score was even below chance level when seen without presentation of audio (VO: 9.52%). However, using relative percentage correct, the success of communication of darkness is above chance (19.15%): listeners' rating of darkness was relatively high for this target timbre compared to when asked to perform the same music with a different timbre.

The results of the percent correct answers for each target timbre are summarized in **Table 2** below, including the specific scores in each audio-visual condition and the average score. These results indicate that percent correct responses for *dark*, *round*, and *bright* timbres were the lowest (dark<round<bright), and percent correct responses for *light*, *relaxed*, and *sharp* timbres were the highest (sharp>relaxed>light). It is noteworthy that the relative percent correct answers regarding round timbre and sharp timbre was considerably higher than the absolute

---

[3]The method of calculating absolute and relative scores of percent correct answers is borrowed from Timmers and Ashley (2007, Experiment 2).
[4]Timbre metaphors will now be italicized for better readability and display of the results.

TABLE 2 | Percentage of listeners' correct answers for a target timbre (values in each AV condition and their average).

Rated timbre	Absolute percentage[a]				Relative percentage[b]			
	Average	**AO**	**AV**	**VO**	**Average**	**AO**	**AV**	**VO**
Dark	**15.34**	12.70	23.81	9.52	**19.15**	17.49	23.33	17.46
Round	**19.15**	13.33	20.63	22.22	**32.98**	28.33	36.51	31.75
Bright	**23.4**	28.57	20.63	21.67	**29.26**	23.81	31.75	31.67
Velvety	**24.34**	23.81	22.22	26.98	**29.63**	26.98	30.16	31.75
Tense	**26.98**	19.05	30.16	31.75	**25.4**	20.63	25.40	30.16
Dry	**27.13**	31.75	26.98	21.67	**23.94**	25.40	22.22	25.00
Heavy	**31.91**	36.67	28.57	30.16	**30.32**	38.33	28.57	25.40
Light	**34.92**	30.16	36.51	38.10	**39.15**	39.68	42.86	34.92
Relaxed	**35.45**	31.75	34.92	39.68	**34.92**	28.57	38.10	38.10
Sharp	**39.68**	47.62	42.86	28.57	**52.38**	61.90	55.56	39.68
Average	**27.83**	27.54	28.73	27.03	**31.71**	31.11	33.45	30.59

[a]When the rating of the target timbre was higher than ratings of the other nine timbres when the listener was presented with the performance of a target timbre.
[b]When the rating of target timbre was relatively higher for the performance with target timbre than in the other performances of a piece.

percentage. This implies that the performed piece did not sound very round or sharp, but performers were able to change the degree to which those excerpts were perceived as round or sharp. When considering the influence of AV stimuli on the absolute/relative percent correct score, *tense* and *relaxed* timbre were more likely to be rated as high when visual information was presented—as can be seen by the higher percent correct score in both absolute and relative percentages in the audio-visual and the visual-only condition compared to the audio-only condition. The impact of the AV stimuli on other timbre evaluations was more variable, the details of which will be investigated in the next analysis of variance of the timbre ratings.

To examine whether the evaluation of each timbre was above chance level, a one-sample *t*-test was conducted, in which the test value was 0.1. This is a chance probability that a target timbre is the highest rated timbre. The values of t, df, and mean score for the one-sample t-test are summarized below (**Table 3**). This analysis showed that the mean values of all variables were significantly higher than the chance level of 10%, which gives the first indication of a reliable communication of the target instruction. Thus, the first hypothesis (H1) was supported. Interestingly, for nine timbres the communication was even successful in the visual-only condition, except for the *dark* timbre (9.52% in VO condition).

Three-Way ANOVA Results

In the repeated measures three-way ANOVA, the main effect of heard timbre (two levels: target timbre, opposite timbre) is a key indicator of successful timbral communication, as the score indicates the differences in timbre ratings across two performances with two contrasting timbres (target, opposite). In other words, significant differences suggest that listeners can successfully recognize performers with target timbres, otherwise performances with target timbre and opposite timbre are confounded. The overall ANOVA results are in **Table 4**.

TABLE 3 | *T*-value, *p*-value, and mean score of absolute percent correct in the one-sample *t*-test.

Rated timbres	Mean	*t*-value	df	*p*-value
Dark percent correct	0.15	2.03	188	<0.05
Round percent correct	0.19	3.18	187	<0.05
Bright percent correct	0.23	4.33	187	<0.01
Velvety percent correct	0.24	4.58	188	<0.01
Tense percent correct	0.27	5.25	188	<0.01
Dry percent correct	0.27	5.27	187	<0.01
Heavy percent correct	0.32	6.43	187	<0.01
Light percent correct	0.35	7.17	188	<0.01
Relaxed percent correct	0.35	7.30	188	<0.01
Sharp percent correct	0.40	8.32	188	<0.01

Main Effects

We found the main effect of heard timbre for all timbre ratings except for the *dark* timbre, indicating that the listeners failed to recognize performances with *dark* timbre regardless of who was performing and in which condition. This result is in line with the results in section Participants, which show that the percentage of correct answers for *dark* evaluation is the lowest (15%) among the 10 timbres. Therefore, H2 was mostly supported - the listeners can successfully differentiate between performances with contrasting timbral intentions except when the music was played with dark timbre.

Two-Way and Three-Way Interaction

Heard Timbre × condition. The two-way interaction between heard timbre and condition (three level: AO, AV, and VO) suggests that the recognition of a target timbre is influenced by auditory-visual presentation modes. We found a significant interaction effect between heard timbre and AV condition for the communication of three timbres: *relaxed*, *tense*, and *velvety*. For *relaxed* timbre evaluation, planned contrast analysis between

TABLE 4 | Three-way repeated measures ANOVA main effects and interactions for listeners' evaluation of 10 timbres.

Effects		Brightness	Darkness	Heaviness	Lightness	Roundness	Sharpness	Relaxed	Tense	Dry	Velvety
						Evaluation					
Timbre	F	**22.88****	3.92	**59.02****	**52.44****	**32.38****	**45.38****	**39.05****	**45.30****	**31.18****	**70.25****
	ηp^2	0.77**	0.2	0.76**	0.73**	0.66**	0.73**	0.66**	0.70**	0.62**	0.79**
	df	(1, 16)	(1, 16)	(1, 19)	(1, 19)	(1, 17)	(1, 17)	(1, 20)	(1, 20)	(1, 19)	(1, 19)
Pianist/ Piece	F	**11.90****	**31.48****	**73.12****	**57.28****	**6.80***	**10.25****	**8.42****	**12.83****	**8.93****	**17.22****
	ηp^2	0.43	0.66**	0.79**	0.75	0.53*	0.61	0.30**	0.39**	0.32**	0.48**
	df	(2, 32)	(2, 32)	(2, 38)	(2, 38)	(2, 34)	(2, 34)	(2, 40)	(2, 40)	(1, 19)	(1, 19)
AV stimuli	F	0.78	**4.32***	**3.81***	0.91	**11.56****	**14.75****	0.48	1.39	1.02	**4.37***
	ηp^2	0.05	0.21*	0.17*	0.05	0.64**	0.68**	0.02	0.07	0.05	0.19*
	df	(2, 32)	(2, 32)	(2, 38)	(2, 38)	(2, 34)	(2, 34)	(2, 40)	(2, 40)	(2. 38)	(2. 38)
Timbre * Piece	F	0.93	0.86	1.28	**7.96****	**11.31****	**8.21****	**8.11****	3.13	2.85	**3.84***
	ηp^2	0.06	0.05	0.06	0.30**	0.63**	0.57**	0.29**	0.14	0.13	0.17*
	df	(2, 32)	(2, 32)	(2, 38)	(2, 38)	(2, 34)	(2, 34)	(2, 40)	(2, 40)	(2. 38)	(2. 38)
Timbre * AV	F	0.28	0.83	0.17	0.11	0.11	1.09	**3.81***	**6.95****	2.29	**3.80***
	ηp^2	0.02	0.05	0.01	0.01	0.01	0.06	0.16*	0.26**	0.11	0.17*
	df	(2, 32)	(2, 32)	(2, 38)	(2, 38)	(2, 34)	(2, 34)	(2, 40)	(2, 40)	(2. 38)	(2. 38)
Piece * AV	F	1.86	1.36	2.21	0.91	2.22	1.55	**3.10***	0.53	0.62	0.41
	ηp^2	0.10	0.08	0.10	0.05	0.12	0.08	0.13*	0.03	0.03	0.02
	df	(4, 64)	(4, 64)	(4, 76)	(4, 76)	(4, 68)	(4, 68)	(4, 80)	(4, 80)	(4, 76)	(4, 76)
Timbre* Piece * AV	F	0.31	1.16	**4.80****	1.20	1.52	**4.04***	1.83	0.73	0.75	**3.31***
	ηp^2	0.02	0.07	0.45**	0.06	0.08	0.19*	0.08	0.04	0.04	0.15*
	df	(4, 64)	(4, 64)	(4, 76)	(4, 76)	(4, 68)	(4, 68)	(4, 80)	(4, 80)	(4, 76)	(4, 76)

**p < 0.01; *p < 0.05.
Bold values indicate the significant results in the main effects and interaction effects.

conditions contains vision and sound-only condition suggested that the communication was better when both audio and vision contained than the sound-only condition, $F_{(1, 20)} = 15.56$, $p < 0.01$, $\eta p^2 = 0.44$. For *tense* timbre evaluation, planned contrast results indicated that tension was differentiated more clearly in the AV conditions than in the audio-only condition, $F_{(1, 20)} = 5.00$, $p < 0.05$, $\eta p^2 = 0.20$. **Figure 4** displays the evaluation for relaxed timbre (left) and tense timbre respectively, where the interval (i.e., differences in ratings across two performance with target timbre and opposite timbre) was larger in visual conditions. However, the participants differentiated the *velvety* timbre more clearly in the audio-only condition than in the audio-visual condition, $F_{(1, 20)} = 9.10$, $p < 0.01$, $\eta p^2 = 0.31$ (**Figure 5**).

Heard Timbre × condition × piece/performer. The three-way interaction between heard timbre, condition, and piece/performer suggests the extent to which a successful communication of a target timbre in particular audio-visual conditions is reliant on the performer/piece. We found significant interaction results for the communication of *heavy* timbre (**Figure 6**), *sharp* timbre (**Figure 7**), and *velvety* timbre (**Figure 8**). The results of planned contrasts were labeled in these figures where significant three-way interaction was found. More specifically, P1 communicated *heavy* timbre better than P3 in aural conditions while P3 communicated better in the visual-only condition; P1 communicated *sharp* timbre better in

the visual-only condition while P3 communicated better in aural conditions. In addition, P1 communicated *velvety* timbre more effectively in aural conditions than P2 whereas P2 was better in the visual-only condition.

To sum up, H3 was partially supported: the above results suggest that only the communication of *relaxed* and *tense* timbres is not reliant on the performer's differences. In other cases, there is an individual difference in the utilization of the audio-visual condition to communicate the timbral intention to the listeners. Each performer may be specialized in communicating a particular timbre with particular aural or visual cues.

Non-verbal Sensory Judgment Analysis

Judgment of the size, shape, and brightness was collected as categorical data in the form of either: A (bigger size, rounder shape, and brighter version), or B (smaller size, sharper shape, and darker version) in the questionnaire. This data was replaced with a score of either 0 or 1, to give a method of calculating the mean across different participants ($N = 21$). The measure of each type of judgment (e.g., size evaluation while hearing *bright* timbre) for each participant was calculated *via* an average across AV stimuli and three music pieces (Mean = SUM divided by 9). **Table 5** displays the value of the size, shape, and brightness evaluations in response to five pairs of timbre intentions. A paired sample *T*-test was conducted to compare the mean difference in non-verbal judgments in response to
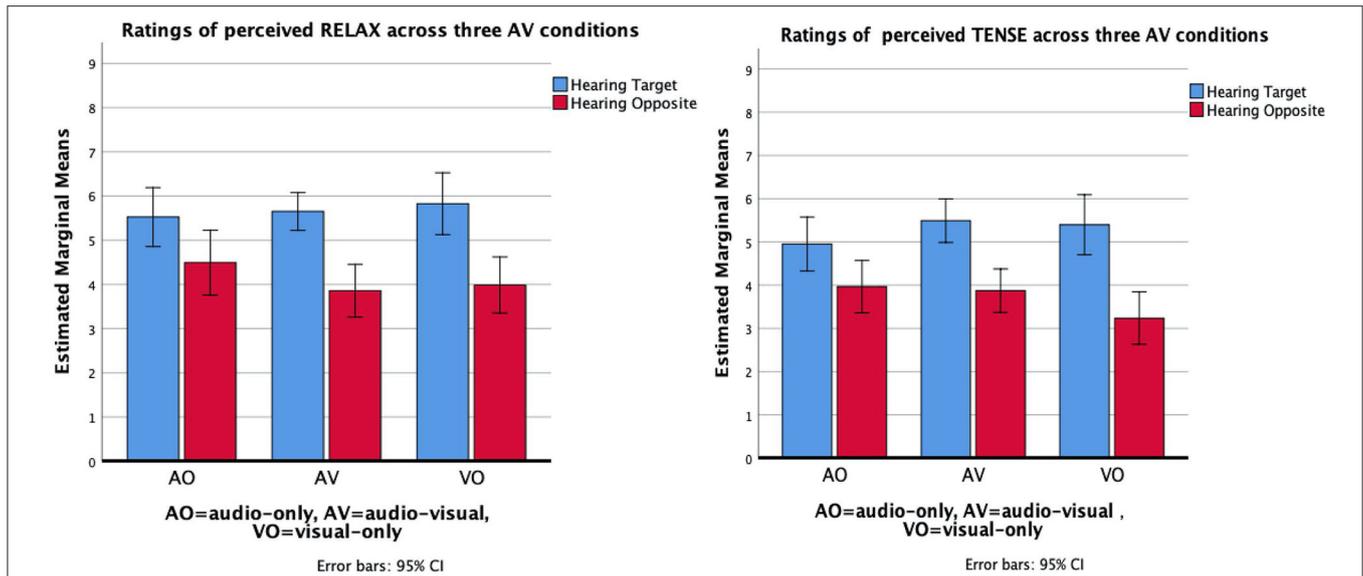
FIGURE 4 | Mean values of the evaluation of *Relaxed* timbre **(left)** and *Tense* timbre **(right)** across three AV stimuli found by averaging the pianist/piece factor. Error bars represent confidence intervals.

contrasting performances—e.g., the mean difference in size evaluation between hearing *bright* timbre and hearing *dark* timbre. Significant differences are shown with *.

The results indicate a strong timbre-shape association in all five pairs of timbral intentions, hence H4 was supported. Participants tended to choose the round object when hearing/seeing performances with *dark, round, light, relaxed,* and *velvety* timbral intentions and chose the sharp object when hearing/seeing performances with *bright, sharp, heavy, tense,* and *dry* timbral intentions. In contrast, size was associated with the perception of two pairs of timbres (*bright/dark* timbre, *heavy/light* timbre), and brightness was associated with only one timbre perception pair (*relaxed/tense* timbre).

Principal Components Analysis (PCA) and Acoustic/Visual Information

The PCA revealed the presence of three components with eigenvalues >1, which account for a total of 69.6% of the variance in evaluations of the 10 timbres, explaining 38.71, 19.2, and 11.68% of the variance respectively. To aid the interpretation of the three components, varimax rotation was used and the results are displayed in **Table 6**. Loading values that were <0.3 were excluded from the table, so blanks in the table indicate where low loading exists.

As a result, the first dimension (named as Round-dimension, eigenvalue: 38.7%) includes the timbre evaluation of round, velvety, relaxed (positive scores) and non-dry (negative score). This dimension is related to touch and movement qualities and is most strongly associated with Round. It is a combination of tactile feelings of roundness and velvetiness, and low dryness. The second component (named as Heavy-dimension) relates to the evaluation of heavy, dark, tense, and sharp. This is a combination of high intensity and negative valence and is associated with haptic sensations of sharpness and tactile
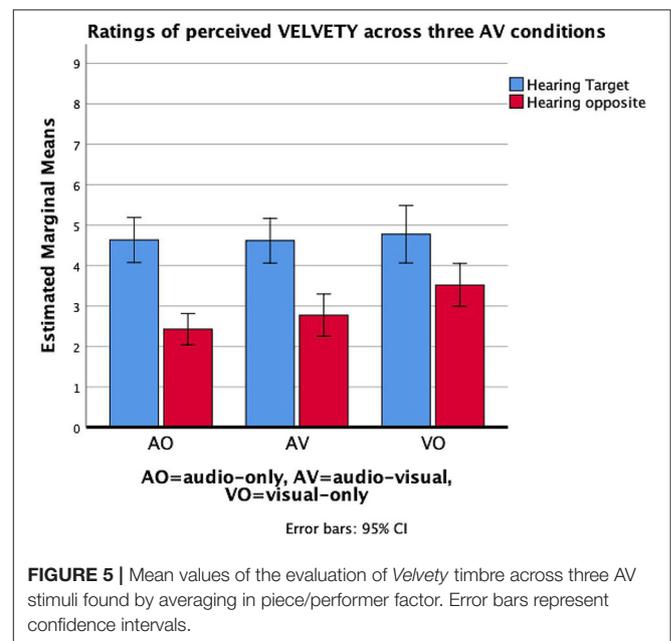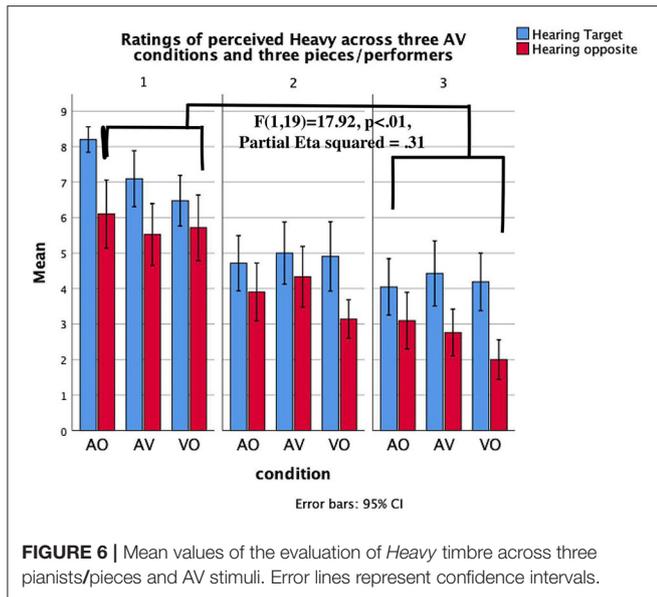

FIGURE 5 | Mean values of the evaluation of *Velvety* timbre across three AV stimuli found by averaging in piece/performer factor. Error bars represent confidence intervals.

sensations of heaviness. It is most strongly associated with Heavy and relates to experiences of weight and negative intensity. The last component (named as Luminance-dimension) is most strongly associated with Brightness followed by lightness. It seems that the Luminance-dimension is both positive in valence, high in space and lightweight.

The **Appendix 4** displays the audio waves and video snapshot[5] of the three pairs of performances (round vs. velvety, heavy

[5]Video snapshots were captured at the same time slot to obtain a relatively fair comparison. However, video shortcuts failed to provide dynamic and continuous gestural changes e.g. how fingers are articulated (smooth or detached).

FIGURE 6 | Mean values of the evaluation of *Heavy* timbre across three pianists/pieces and AV stimuli. Error lines represent confidence intervals.
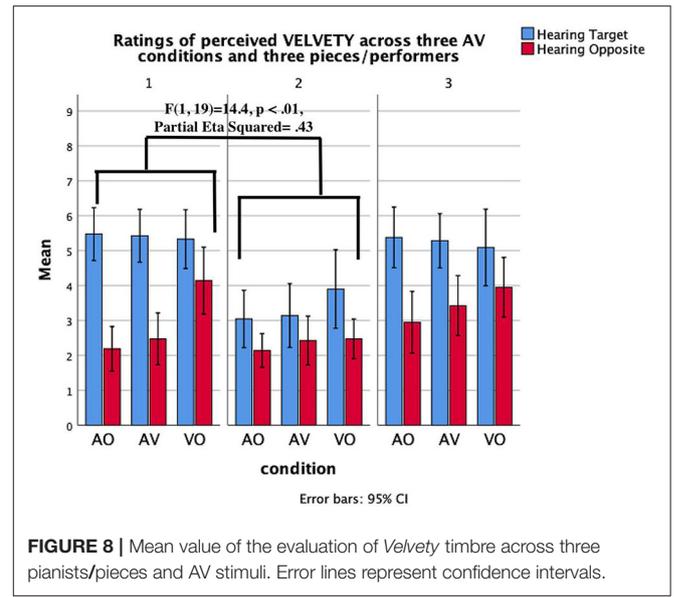
FIGURE 8 | Mean value of the evaluation of *Velvety* timbre across three pianists/pieces and AV stimuli. Error lines represent confidence intervals.
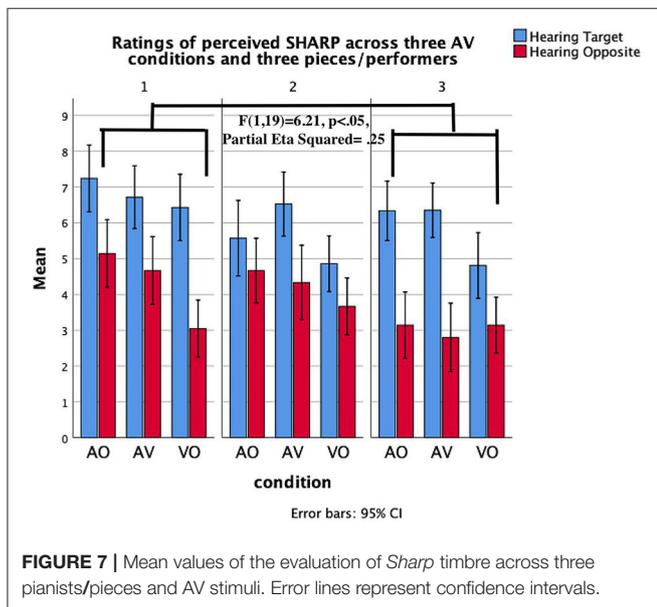
FIGURE 7 | Mean values of the evaluation of *Sharp* timbre across three pianists/pieces and AV stimuli. Error lines represent confidence intervals.

vs. dark, bright vs. light) by three performers. We compared the aural/visual information in pairs, to explore whether the pianists were defaulting to the same performance style/gestures or not when playing these highly correlated timbres. Observable differences were found in these pairs:

Round vs. Velvety: P1 differentiated these two timbres using intensity (*round* is louder than *velvety*); while P2 used performance tempo to make contrast (*round* is quicker than *velvety*). In the visual cues, P1 and P2 changed sound-producing gestures, with higher hand position and curved finger to produce *round* timbre while using lower hand position and flatter finger to play *velvety* timbre. The difference in P3 was little.

Heavy vs. Dark: For P1 and P3, *heavy* timbre was interpreted as much louder than *dark* timbre. P2 played *slower* in dark timbre compared to *heavy* timbre. In general, all the three pianists were shown more solemn facial expression, and P3 showed more intensive feelings with inward upper-body movements toward the piano.

Bright vs. Light: In general, these two timbres were played much softer for the three performers. Differences were found in the facial expression and posture for P1 and P3, that *bright* timbre was played with more delighted and cheerful expression and movements.

DISCUSSION

The Reliability of Communicating Timbre

The first research question examined in this study is whether the communication of timbral intention between pianists and listeners is reliable. The study results indicate that all the 10 timbres are communicated with an accuracy above chance level, although the average percentage correct was considerably higher for some timbres (*sharp, relaxed, and light*) than for others. Furthermore, the repeated-measures ANOVA results showed that nine of 10 timbres were reliably communicated, except for the *dark* timbre. This suggests that timbral intention can be communicated between the performer and the listeners like other abstract intentions including expressivity (Davidson, 1993), emotions (Juslin, 2000), and sensorimotor feelings (De Poli et al., 2017). There is already evidence demonstrating that instrument-expertise influences the perception of musical expressivity due to the activation of sensory representation in the observation of motor plan (Broughton and Davidson, 2014).

The difficulty of communicating the 10 timbres varies one from the other. *Bright* timbre was the easiest to communicate and not influenced by the AV condition factor or performer/piece, while *dark* timbre was the most difficult one and the

TABLE 5 | Mean judgment scores for size, shape, and brightness in the perception of 10 timbres.

	Size	Shape	Brightness
	(small) 0 – 1 (big)	(sharp) 0 – 1 (round)	(dark) 0 – 1 (bright)
Bright	0.47*	0.45**	0.53
Dark	0.35*	0.62**	0.57
Round	0.49	0.66**	0.54
Sharp	0.59	0.17**	0.44
Heavy	0.58**	0.41**	0.53
Light	0.31**	0.72**	0.60
Tense	0.41	0.40**	0.48*
Relaxed	0.38	0.68**	0.63*
Dry	0.39	0.38**	0.54
Velvety	0.46	0.76**	0.57

**p < 0.01; *p < 0.05.*

TABLE 6 | Factor loading after varimax rotation.

Evaluation	Component 1 Round (Touch-movement)	Component 2 Heavy (Weight-Negative intensity)	Component 3 Bright (Luminance)
Round	**0.851**		
Velvety	**0.827**		
Dry	−0.626	0.333	0.340
Relaxed	0.595		0.486
Heavy		**0.811**	−0.328
Dark		**0.765**	−0.320
Tense	−0.473	0.650	
Sharp	−0.595	0.614	
Bright			**0.829**
Light		−0.344	**0.744**

Bold values highlight the most two representative timbres in each dimension.
The highlighted red mark indicates the three clusters in PCA. They were framed for readability and highlighting.

communication even failed in the visual-alone condition. One of the potential reasons could be that this experiment uses a Yamaha piano which usually has a brighter sound, resulting in the superior communication of brightness.

The differences in the communication outcome imply that music performers may consider suitable strategies (i.e., acoustic cues or visual cues) for the expression and communication of different timbres and that music educators are encouraged to focus on the teaching and learning of more difficult ones (i.e., *dark, round, and bright*) in piano lessons. Facial expression can be considered to differentiate *bright* and *dark* timbre, which has been demonstrated to be an effective way of differentiating the effects conveyed in the vocal performance (Thompson et al., 2005). For example, **Figure 9** below shows that two performers in this research used different visual communication strategies to differentiate *bright* and *dark* timbre: one used facial expression while the other one used forward/inward upper-body movements.
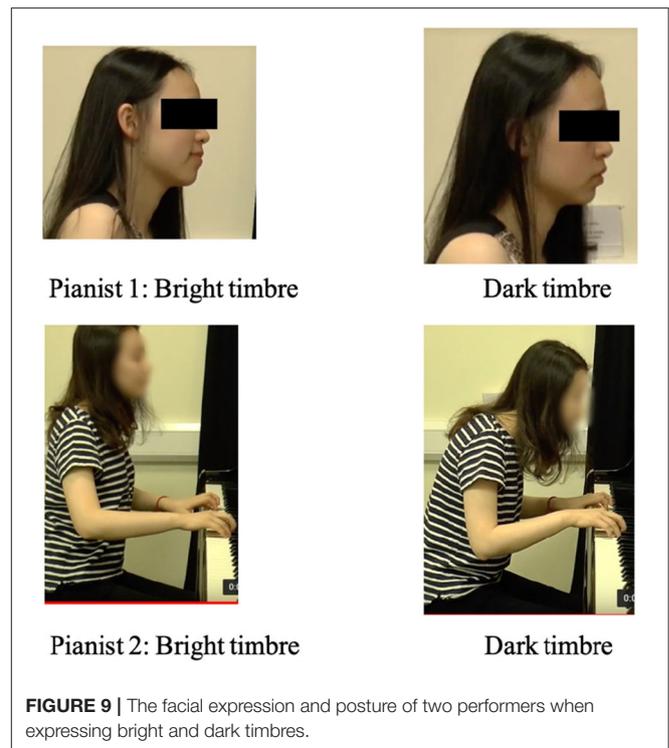
FIGURE 9 | The facial expression and posture of two performers when expressing bright and dark timbres.

Hearing and Seeing Piano Timbre

The subsequent research question we investigated concerns the relevance of auditory and/or visual cues in the communication of timbral intention. Firstly, the comparison of sound waves across five pairs of timbres indicated that the inconsistency in the choice of other performance parameters (either using intensity or performance tempo) among the three pianists; however, the reliability of communicative process implied that the participants had a common, consensual understanding of the meaningful use of performance parameters to achieve certain timbre. This result is in line with Bernays and Traube's (2014) study that showed an individuality in the playing style of five pianists in the expression of the same timbral intention, with a personalized choice of musical features (articulation, pedal, attack, dynamics, etc.).

Secondly, the importance of visual cues in music performance is verified for the paring of *relaxed* and *tense* timbres, where visual information determined the effectiveness of timbral communication independent of the pianist/piece. For the remaining timbres (*heavy, light, round, sharp, dry, and velvety*), the influence of the AV condition was in most cases accompanied by an interaction involving the effect of pianist/piece (e.g., when communicating *sharp* timbre, P1 is better in visions while P3 is better in sounds), implying the difference in the preference of using visual cues among the pianists. Interestingly, increasing numbers of scholars (Thompson et al., 2005; Schutz and Lipscomb, 2007; Schutz, 2008; Davidson, 2012) have started to claim the benefits of using bodily communication in a music performance for the performers, including enhancing auditory experience, shaping the perception of musical notes, or facilitating the comprehension of lyrics. According to Schutz (2008), percussive instruments such as

the piano and marimba could be most successful in benefiting from visual communication due to limited timbral nuances and percussive/short note length. We therefore encourage pianists to use bodily communication (i.e., appropriate facial expressions, gestures, and postures) in the communication of timbral intention, in addition to the auditory component of the music performance.

We integrated two angles including sound-producing gestures and ancillary gestures in the video stimuli. Although it is unknown whether the visual communication of 10 timbres differs in the reliance on specific aspects of bodily movements, we anticipate that the visual information of sound-producing gestures underlies the perception of 10 timbres—seeing how the keyboard is touched (i.e., percussive/non-percussive, curved/flat, hard/soft, etc.) influenced the listeners' perception of piano timbre, as suggested by the McGurk effect in music perception (Saldaña and Rosenblum, 1993; Schutz and Lipscomb, 2007). Further study needs to be taken to investigate the impact of ancillary gestures on timbre perception.

Last but not the least, the successful communication of timbral intention even in the vision-alone condition suggests an embodied perspective in music communication. One possible explanation is that understanding a pianist's silent movements possibly involves the simulation of a performative action in the listener's head, which leads to the triggering of an auditory image of the timbral effect (Keller, 2012). This is in line with previous studies (Camurri et al., 2003; Dahl and Friberg, 2007) showing that listeners can still detect the emotional intention of performers even when the sound information is unavailable. Brain imaging studies supported this and found that there are activations of certain motor-related areas in the brain when imagining music with the "inner ear" (Zatorre and Halpern, 2005) and in the perception of musical sounds (Haueisen and Knösche, 2001). This study extends the view of embodied music listening, by showing the successful communication of timbral intentions even without aural cues.

In this study, the listeners can "stand in someone else's shoes" to understand the actions, intentions, and sounds of the pianists by simulating the motor activity. According to Molnar-Szakacs and Overy (2006), a "similar or equivalent motor network is engaged by someone listening to singing/drumming as the motor network engaged by the actual singer/drummer" (p. 236) and therefore the communication of timbre was also successful in audio-alone conditions due to a simulation of a motor activity similar or equivalent to that associated with the timbral effect. The loop between sensory experience and motor commands has been explained by the "internal model" in the research into sensorimotor perceptions (Keller, 2012; Maes et al., 2014), which suggests its benefits for musicians themselves including action planning and self-monitoring (Novembre and Keller, 2014) and the benefits for co-players such as synchronization and coordination (Keller et al., 2007). This study demonstrated the benefits of an internal model in performer-listener communication, which is a shared understanding of actions and sounds that occur in the pianists' communicative process of timbral intentions to the listeners. We, therefore, suggest that the listeners are also performers because a mirror

process happens in the perception of piano timbre through the encoding of expressive gestures into sounds (timbres) and the decoding of sounds (timbres) into expressive gestures (cf. Leman and Maes, 2014).

Timbre-Related Cross-Modal Correspondence

The results of non-verbal sensory judgement indicated that a cross-modal timbre-shape association was found in all 10 timbres. When music was played with *dark*, *round*, *light*, *relaxed*, and *velvety* timbral intentions, it was felt to be rounder than when played with *bright*, *sharp*, *heavy*, *tense,* and *dry* intentions. This timbre-shape association is in line with several previous studies which have found a sharpness/color association with either *soft* or *harsh* timbres (Adeli et al., 2014) as well as visual textural associations (Giannakis, 2006). In addition to a timbre-shape association, this study also found a timbre-size association in two pairs of timbres (*bright/dark*, *heavy/light*).

Wallmark and Kendall (2018) applied the theory of conceptual metaphor (Lakoff and Johnson, 1980) to understand timbre semantics. They explained that when timbre is described with light contrast (e.g., bright/dark) and textural feature (e.g., rough, smooth), it reflects the conceptual metaphors that SOUND IS LIGHT and SOUND IS TEXTURE as people use the source domain of vision and tactile sensations as a reference to drive the meaning of the target domain (i.e., timbre). This statement reflects how abstract domains of human experience can be understood by concrete, embodied accounts. The embodied accounts to explain timbre-related CMC can also be found in Spence (2011) who provided three for the underlying reasoning of CMC: statistical, semantic-mediated, and embodied accounts. Statistical and embodied accounts explained the results of repeated exposure to, and physical interaction with, instruments and the extent to which the weak synaesthesia experience related to piano timbre can be established as a part of musical training (Spence, 2011). Studies have indicated that repeated exposure to statistical co-current pairs of stimuli can help to shape the coupling pairs, even in circumstances where the stimuli are unrelated [e.g., the association between a feeling of stiffness and visual luminance, Ernst (2007)]. Semantic-mediated accounts help to understand the linguistic descriptions related to piano timbre and their possible impact in generating a cross-modal coupling experience [cf. Dolscheid et al. (2013)]. For example, in the production of round timbre, a teacher's verbal description (round), modeling, and the embodiments (round handshape) may help the student to establish an association between the sound outcome and the round shape.

Piano Timbre as a Component of Expressivity

For a long time, the definition of musical expression/expressiveness has been deeply influenced by Seashore's 1938 statement on "deviation from the score"[6] [see a discussion in Clarke and Doffman (2014)]. The shortcomings of

---

[6]which defines expression as a deviation from the notated values (i.e. pure tone, true pitch, even dynamics, rigid rhythm etc.) on the musical score.

this definition and the over-emphasis on the sonic properties of performance have been noted by several scholars, for instance leading to a trend of the musical score being the primary ontological focus of music (Dogantan-Dack, 2014) and a disembodied, ahistorical account of musical performance by conceptualizing the score as the piece (Clarke, 2004). This research would like to assure that, in the pursuit of timbral nuances, musical expressiveness is imprinted in the acoustic variance and the subtle control of gestures and touch of pianists, as if "touch is the expressive skeleton on which the pianist enfolds the expressive flesh" (Dogantan-Dack, 2014, p. 7). Touch, gestures, bodily movements in piano playing actively shape a listener's multimodal perception of piano timbre, resulting in sensorimotor perception and cross-modal correspondences.

IMPLICATIONS AND CONCLUSIONS

There are some implications for instrumental music pedagogy and performance. This study offers interesting insights into the question of "what do listeners perceive in timbre in music performance." From the listener's perspective, the perception of timbre is not merely variations in acoustic information but related to multimodal perception including kinaesthetic and muscular sensations (heavy-light, tensed-relax) and cross-modal correspondences (physical size and shape). This is consistent with the studies of the semantics of timbre (Bellemare and Traube, 2005; Bernays, 2013; Kojucharov and Rodà, 2015; Saitis and Weinzierl, 2019) that have shown rich subjective experiences and multidimensionality in response to timbre. Therefore, music teachers and students are encouraged to use cross-domain metaphors and multimodal communications (gestures, modeling, touch, etc.) in the teaching and learning of timbre-related performance goals in music lessons (Li and Timmers, 2021).

This study suggests that timbral intention can work well as a bridge that connects the performer and the listener, due to the resonance of sensorimotor knowledge induced by musical sounds. It also has implications for music education that teachers should encourage students to mobilize timbre as a deeper motivation for expressing intention/message and communicating it to their listeners. As Dogantan-Dack (2017) pointed out, classical musicians are facing pressure due to the sustainability of the profession itself and the diversity of today's musical genres and practices. In the process of working on timbral intentions and communicating them to the listeners, student-pianists may increase their musical competency and sense of autonomy [i.e., expressive freedom extending beyond the composer's intention, Dogantan-Dack (2017)], which may help to fulfill their psychological needs at this stage of their musical education and maintain the motivation to persist and engage with music (Wise et al., 2017).

One of the limitations in this study is the unsystematic control of performance parameters such as dynamics and tempo, and we acknowledge that these elements can influence musical perception. This enables a more open musical interpretation for the performers to express different timbral intentions,

which probably work as co-variables to influence the listeners' judgments. Another limitation is the limited sample size used in the present study; thus, the results of the variety in communication outcomes of 10 timbres must be taken as tentative. Future research on the visual communication of piano timbre may consider using less timbres, for instance relaxed/tense and heavy/light, to expand our understanding of touch qualities (e.g., soft/hard, percussive/non-percussive) with deeper insights into bodily feelings in piano playing. Alternatively, future research may also consider using non-pianists as the control group (i.e., the level of musical training as the between-group factor). There is already evidence demonstrating that instrument-expertise influences the perception of musical expressivity due to the activation of sensory representation in the observation of motor plan (Broughton and Davidson, 2014). Future research can consider more restricted control of performance parameters using a single piano tone or chord. Furthermore, future research may be able to use point-light displays to replace bodily movement in the videos, to avoid the influence of familiarity of performers or their facial expressions. This method could also help to create congruent or incongruent stimuli, by synchronizing the movement features of one stimulus with the performance data of another in terms of onset/offset time, duration, dynamics, and pedaling [cf. Vuoskoski et al. (2014)]. For example, a new artificial video can be generated *via* synchronizing movement features in a "tense" performance with the sound signal of a "relaxed" performance thereby allowing an examination of the extent to which visual information influences or modifies listeners' judgments.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Sheffield. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the participants for the publication of any identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SL and RT designed the experiment. SL was responsible for data collection, data analysis, and the draft of manuscript. RT and WW contributed to refine arguments, the presentation of results, and to improve the readability. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717842/full#supplementary-material

Supplementary Video 1 | P1_bright.

Supplementary Video 2 | P1_dark.

Supplementary Video 3 | P1_heavy.

Supplementary Video 4 | P1_light.

Supplementary Video 5 | P1_relax.

Supplementary Video 6 | P1_tense.

Supplementary Video 7 | P1_round.

Supplementary Video 8 | P1_sharp.

Supplementary Video 9 | P1_dry.

Supplementary Video 10 | P1_velvety.

REFERENCES

Adeli, M., Rouat, J., and Molotchnikoff, S. (2014). Audiovisual correspondence between musical timbre and visual shapes. *Front. Hum. Neurosci.* 8:352. doi: 10.3389/fnhum.2014.00352

Balkwill, L. L., and Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Percept.* 17, 43–64. doi: 10.2307/40285811

Baraldi, F. B., Poli, G. D., and Rodá, A. (2006). Communicating expressive intentions with a single piano note. *J. New Music Res.* 35, 197–210. doi: 10.1080/09298210601045575

Barthet, M., Depalle, P., Kronland-Martinet, R., and Ystad, S. (2010). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Percept.* 28, 135–154. doi: 10.1525/mp.2010.28.2.135

Behne, K. E., and Wöllner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Music. Sci.* 15, 324–342.

Bellemare, M., and Traube, C. (2005). "Verbal description of piano timbre: exploring performer-dependent dimensions," in *Proceedings of the 2nd Conference on Interdisciplinary Musicology (CIM05)*, Montreal, QC: CIM.

Bernays, M. (2013). *The Expression and Production of Piano Timbre: Gestural Control and Technique, Perception and Verbalisation in the Context of Piano Performance and Practice.* (Doctoral dissertation), Université de Montréal, Montreal, QC, Canada.

Bernays, M., and Traube, C. (2014). Investigating pianists' individuality in the performance of five timbral nuances through patterns of articulation, touch, dynamics, and pedaling. *Front. Psychol.* 2014:157. doi: 10.3389/fpsyg.2014.00157

Broughton, M., and Stevens, C. (2009). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychol. Music* 37, 137–153. doi: 10.1177/0305735608004511

Broughton, M. C., and Davidson, J. W. (2014). Action and familiarity effects on self and other expert musicians' Laban effort-shape analyses of expressive bodily behaviors in instrumental music performance: a case study approach. *Front. Psychol.* 5:1201. doi: 10.3389/fpsyg.2014.01201

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments.* Berkeley, Los Angeles, CA: University of California Press. doi: 10.1525/9780520350519

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *Int. J. Hum. Comp. Stud.* 59, 213–225. doi: 10.1016/S1071-5819(03)00050-8

Clarke, E. (2004). "Empirical methods in the study of performance," in *Empirical Musicology: Aims, Methods, Prospects,* eds E. Clarke and N. Cook (Oxford: Oxford University Press), 77–102. doi: 10.1093/acprof:oso/9780195167498.003.0005

Clarke, E., and Doffman, M. (2014). "Expressive performance in contemporary concert music," in *Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures,* eds D. Fabian, R. Timmers and E. Schubert (Oxford: Oxford University Press), 98–114. doi: 10.1093/acprof:oso/9780199659647.003.0006

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Edn.).* Hillsdale, NJ: Erlbaum.

Cox, A. (2016). *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking.* Bloomington, IN; Indianapolis, IN: Indiana University Press. doi: 10.2307/j.ctt200610s

Dahl, S., and Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Percept.* 24, 433–454. doi: 10.1525/mp.2007.24.5.433

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychol. Music* 21, 103–113. doi: 10.1177/030573569302100201

Davidson, J. W. (1995). "What does the visual information contained in music performances offer the observer? Some preliminary thoughts," in *Music and the Mind Machine*, ed R. Steinberg (Berlin, Heidelberg: Springer), 105–113. doi: 10.1007/978-3-642-79327-1_11

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: two distinctive case studies. *Psychol. Music* 40, 595–633. doi: 10.1177/0305735612449896

Davidson, J. W., and Correia, J. S. (2002). "Body movement," in *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*, eds R. Parncutt and G. E. McPherson (Oxford; New York, NY: Oxford University Press), 237–252.

De Poli, G., Mion, L., and Rodá, A. (2009). Toward an action based metaphor for gestural interaction with musical contents. *J. New Music Res.* 38, 295–307. doi: 10.1080/09298210902773941

De Poli, G., Murari, M., Canazza, S., Rodà, A., and Schubert, E. (2017). "Beyond emotion: multi-sensory responses to musical expression," in *The Routledge Companion to Embodied Music Interaction*, eds M. Lesaffre, P. J. Maes, and M. Leman (New York, NY; London: Routledge), 78–86. doi: 10.4324/9781315621364-9

De Poli, G., Rodà, A., and Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *J. New Music Res.* 27, 293–321. doi: 10.1080/09298219808570750

Dogantan-Dack, M. (2014). "Philosophical reflections on expressive music performance," in *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures*, eds D. Fabian, R.

Timmers and E. Schubert (Oxford: Oxford University Press), 3–21. doi: 10.1093/acprof:oso/9780199659647.003.0001

Dogantan-Dack, M. (2017). "Expressive freedom in classical performance: insights from a pianist-research," in *Musicians in the Making: Pathways to Creative Performance*, eds J. Rink, H. Gaunt, and A. Williamon (Oxford: Oxford University Press), 131–135.

Dolscheid, S., Shayan, S., Majid, A., and Casasanto, D. (2013). The thickness of musical pitch: psychophysical evidence for linguistic relativity. *Psychol. Sci.* 24, 613–621. doi: 10.1177/0956797612457374

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 7–7. doi: 10.1167/7.5.7

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Method.* 39, 175–191. doi: 10.3758/BF03193146

Faure, A. (2000). *Des Sons Aux Mots, Comment Parle-t-on du Timbre Musical?* (Doctoral dissertation), Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

Gabrielsson, A., and Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007

Giannakis, K. (2006). A comparative evaluation of auditory-visual mappings for sound visualisation. *Org. Sound* 11, 297–307. doi: 10.1017/S1355771806001531

Godøy, R. I. (2006). Gestural-sonorous objects: embodied extensions of Schaeffer's conceptual apparatus. *Org. Sound* 11, 149–157. doi: 10.1017/S1355771806001439

Goebl, W., Bresin, R., and Fujinaga, I. (2014). Perception of touch quality in piano tones. *J. Acoust. Soc. Am.* 136, 2839–2850. doi: 10.1121/1.4896461

Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., and Warren, J. D. (2009). It's not what you play, it's how you play it: timbre affects perception of emotion in music. *Quart. J. Exp. Psychol.* 62, 2141–2155. doi: 10.1080/17470210902765957

Haueisen, J., and Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792. doi: 10.1162/08989290152541449

Holmes, P. A. (2011). An exploration of musical communication through expressive use of timbre: the performer's perspective. *Psychol. Music* 40, 301–323. doi: 10.1177/0305735610388898

Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and Leman, M. (2010). "Concepts and methods in research on music-related gestures," in *Musical Gestures: Sound, Movement, and Meaning*, eds R. I. Godøy and M. Leman (New York, NY; London: Routledge), 12–35.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. London: University of Chicago Press. doi: 10.7208/chicago/9780226177847.001.0001

Juslin, P. N. (1998). *A Functionalist Perspective on Emotional Communication in Music Performance. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences, 78*. Uppsala: Uppsala University Library.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1797–1813. doi: 10.1037/0096-1523.26.6.1797

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bullet.* 129:770. doi: 10.1037/0033-2909.129.5.770

Juslin, P. N., and Lindström, E. (2010). Musical expression of emotions: modelling listeners' judgements of composed and performed features. *Music Anal.* 29, 334–364. doi: 10.1111/j.1468-2249.2011.00323.x

Juslin, P. N., and Timmers, R. (2010). "Expression and communication of emotion in music performance," in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and S. A. John (New York, NY: Oxford University Press), 453–492.

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Ann. N. Y. Acad. Sci.* 1252, 206–213. doi: 10.1111/j.1749-6632.2011.06439.x

Keller, P. E., Knoblich, G., and Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action

simulation in synchronization. *Consciousness Cogn.* 16, 102–111. doi: 10.1016/j.concog.2005.12.004

Kendall, R. A., and Carterette, E. C. (1990). The communication of musical expression. *Music Percept.* 8, 129–163. doi: 10.2307/40285493

Kochevitsky, G. (1967). *The Art of Piano Playing: A Scientific Approach*. Secaucus, NJ: Summy-Birchard Music.

Kojucharov, E., and Rodà, A. (2015). "Towards an empirically based definition of piano touch: touch/timbre relationship and key motion measurements," in *Proceedings of the 9th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM2015)*, eds J. Ginsborg, A. Lamont, and S. Bramley. (Manchester, England), 516–521.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Leman, M. (2008). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7476.001.0001

Leman, M., and Maes, P. J. (2014). The role of embodiment in the perception of music. *Empir. Musicol. Rev.* 9, 236–246. doi: 10.18061/emr.v9i3-4.4498

Li, S., and Timmers, R. (2020). Exploring pianists' embodied concepts of piano timbre: an interview study. *J. New Music Res.* 49, 477–492. doi: 10.1080/09298215.2020.1826532

Li, S., and Timmers, R. (2021). Teaching and learning of piano timbre through teacher-student interactions in lessons. *Front. Psychol.* 2021:576056. doi: 10.3389/fpsyg.2021.576056

Maes, P. J., Leman, M., Palmer, C., and Wanderley, M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008

McAdams, S. (2013). "Musical timbre perception," in *The Psychology of Music, 3rd Edn*, ed D. Deutsch (London, Waltham, and San Diego: Academic Press), 35–67. doi: 10.1016/B978-0-12-381460-9.00002-X

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Molnar-Szakacs, I., and Overy, K. (2006). Music and mirror neurons: from motion to'e'motion. *Soc. Cogn. Affect. Neurosci.* 1, 235–241. doi: 10.1093/scan/nsl029

Morimoto, Y., and Timmers, R. (2012). "The effect of repeated listening on pleasure and boredom responses," in *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music,* eds E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastiadis (Thessaloniki), 693–697.

Namba, S., Kuwano, S., Hashimoto, T., Berglund, B., Da Rui, Z., Schick, A., et al. (1991). Verbal expression of emotional impression of sound: A cross-cultural study. *J. Acoust. Soc. Jpn. (E)* 12, 19–29.

Novembre, G., and Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Front. Hum. Neurosci.* 8:603. doi: 10.3389/fnhum.2014.00603

Ortmann, O. (1935). What is tone-quality? *Musical Quart.* 21, 442–450. doi: 10.1093/mq/XXI.4.442

Rozé, J., Aramaki, M., Kronland-Martinet, R., and Ystad, S. (2017). Exploring the perceived harshness of cello sounds by morphing and synthesis techniques. *J. Acoust. Soc. Am.* 141, 2121–2136. doi: 10.1121/1.4978522

Saitis, C., and Weinzierl, S. (2019). "The semantics of timbre," in *Timbre: Acoustics, Perception, and Cognition*, eds K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay (Cham: Springer), 119–150. doi: 10.1007/978-3-030-14832-4_5

Saldaña, H. M., and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Percept. Psychophys.* 54, 406–416. doi: 10.3758/BF03205276

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empir. Musicol. Rev.* 3, 83–108. doi: 10.18061/1811/34098

Schutz, M., and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception* 36, 888–897. doi: 10.1068/p5635

Seashore, C. E. (1938). *The Psychology of Music*. New York, NY: McGraw Hill. doi: 10.2307/3385515

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attent. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7

Thompson, W. F., Graham, P., and Russo, F. A. (2005). Seeing music performance: visual in?uences on perception and experience. *Semiotica* 2005, 203–227. doi: 10.1515/semi.2005.2005.156.203

Thompson, W. F., and Russo, F. A. (2007). Facing the music. *Psychol. Sci.* 18, 756–757. doi: 10.1111/j.1467-9280.2007.01973.x

Timmers, R., and Ashley, R. (2007). Emotional ornamentation in performances of a *Handel* sonata. *Music Percept. Interdiscipl. J.* 25, 117–134. doi: 10.1525/mp.2007.25.2.117

Timmers, R., Marolt, M., Camurri, A., and Volpe, G. (2006). Listeners' emotional engagement with performances of a Scriabin étude: an explorative case study. *Psychol. Music* 34, 481–510. doi: 10.1177/0305735606067165

Van Zijl, A. G., and Sloboda, J. (2011). Performers' experienced emotions in the construction of expressive musical performance: an exploratory investigation. *Psychol. Music* 39, 196–219. doi: 10.1177/0305735610373563

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* 101, 80–113.

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., and Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attent. Percept. Psychophys.* 76, 591–604. doi: 10.3758/s13414-013-0582-2

Wallmark, Z. (2019a). A corpus analysis of timbre semantics in orchestration treatises. *Psychol. Music* 47, 585–605. doi: 10.1177/0305735618768102

Wallmark, Z. (2019b). Semantic crosstalk in timbre perception. *Music Sci.* 2, 1–18. doi: 10.1177/2059204319846617

Wallmark, Z., and Kendall, R. A. (2018). "Describing sound: the cognitive linguistics of timbre," in *The Oxford Handbook of Timbre*, eds E. I. Dolan and A. Rehding (New York, NY: Oxford University Press), 9780190637224.013.14. doi: 10.1093/oxfordhb/9780190637224.013.14

Wanderley, M., and Vines, B. (2006). "Origins and functions of clarinettists' ancillary gestures," in *Music and Gesture,* eds A. Gritten and E. King (Aldershot, UK: Ashgate),165–191.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *J. New Music Res.* 34, 97–113.

Wapnick, J., Mazza, J. K., and Darrow, A. A. (2000). Effects of performer attractiveness, stage behavior, and dress on evaluation of children's piano performances. *J. Res. Music Educ.* 48, 323–335. doi: 10.2307/3345367

Wapnick, J., Ryan, C., Lacaille, N., and Darrow, A. A. (2004). Effects of selected variables on musicians' ratings of high-level piano performances. *Int. J. Music Educ.* 22, 7–20. doi: 10.1177/0255761404042371

Wise, K., James, M., and Rink, J. (2017). "Performers in the practice room," in *Musicians in the Making: Pathways to Creative Performance*, eds J. Rink, H. Gaunt, and A. Williamon (Oxford: Oxford University Press), 143–163. doi: 10.1093/acprof:oso/9780199346677.003.0011

Wöllner, C. (2020). Call and response: musical and bodily interactions in jazz improvisation duos. *Musicae Scientiae* 24, 44–59. doi: 10.1177/1029864918772004

Ystad, S., Aramaki, M., and Kronland-Martinet, R. (2019). "Timbre from sound synthesis and high-level control perspectives," in *Timbre: Acoustics, Perception, and Cognition*, eds K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay (Cham: Springer), 361–389. doi: 10.1007/978-3-030-14832-4_13

Zatorre, R. J., and Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. *Neuron* 47, 9–12. doi: 10.1016/j.neuron.2005.06.013