JULY 03 2024

Acoustic cues of keyboard mechanics enable auditory localization of upright piano tones

Federico Fontana ; Hanna Järveläinen ; Stefano Papetti ; Yuri De Pra

J. Acoust. Soc. Am. 156, 164–175 (2024)
https://doi.org/10.1121/10.0026484

JASA ARTICLE

Acoustic cues of keyboard mechanics enable auditory localization of upright piano tones

Federico Fontana,[1,a)] Hanna Järveläinen,[2] Stefano Papetti,[2] and Yuri De Pra[3]
[1]Department of Mathematics, Computer Science and Physics, Università di Udine, 206 via delle Scienze, Udine 33100, Italy
[2]Institute for Computer Music and Sound Technology, Zürcher Hochschule der Künste, 96 Pfingstweidstrasse, Zurich 8048, Switzerland
[3]Department of Energy Technologies and Renewable Sources, Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile, 301 via Anguillarese, Rome 00123, Italy

ABSTRACT:
Piano tone localization at the performer's listening point is a multisensory process involving audition, vision, and upper limb proprioception. The consequent representation of the auditory scene, especially in experienced pianists, is likely also influenced by their memory about the instrument keyboard. Disambiguating such components is not obvious, and first requires an analysis of the acoustic tone localization process to assess the role of auditory feedback in forming this scene. This analysis is complicated by the acoustic behavior of the piano, which does not guarantee the activation of the auditory precedence effect during a tone attack, nor can it provide robust interaural differences during the subsequent free evolution of the sound. In a tone localization task using a Disklavier upright piano (which can be operated remotely and configured to have its hammers hit a damper instead of producing a tone), twenty-three expert musicians, including pianists, successfully recognized the angular position of seven evenly distributed notes across the keyboard. The experiment involved listening to either full piano tones or just the key mechanical noise, with no additional feedback from other senses. This result suggests that the key mechanical noise alone activated the localization process without support from vision and/or limb proprioception. Since the same noise is present in the onset of the full tones, the key mechanics of our piano created a touch precursor in such tones that may be responsible of their correct angular localization by means of the auditory precedence effect. However, the significance of pitch cues arriving at a listener after the touch precursor was not measured when full tones were presented. As these cues characterize a note and, hence, the corresponding key position comprehensively, an open question remains regarding the contribution of pianists' spatial memory of the instrument keyboard to tone localization.
 https://doi.org/10.1121/10.0026484

(Received 28 December 2023; revised 23 April 2024; accepted 13 June 2024; published online 3 July 2024)
[Editor: Vasileios Chatziioannou] Pages: 164–175

I. INTRODUCTION

Keyboard instruments establish a one-dimensional map, associating each note to a coordinate point belonging to an ideal horizontal axis in front of the player. Concerning acoustic pianos, the hypothesis of an auditory scene for musical pitch finds its physical foundation in the alignment between each key and the respective strings responsible for producing the corresponding tone (Fletcher and Rossing, 1991). Pianists, therefore, internalize this key-to-strings association through repeated practice with their instrument.

Determining whether a spatial correspondence exists between the position of a key and the tone it produces within a pianist's personal space raises complex issues. Is the pianist's auditory horizon populated by tones that move from left to right as their pitch increases? Does the auditory localization of tones establish a spatial correspondence with the strings generating them? Answering these questions is far from trivial, as they involve multiple senses, perception, and cognition. When a pianist intentionally presses a key, several sensory channels come into play simultaneously: They not only hear the corresponding tone but also receive tactile precursor feedback, proprioceptive information, and occasionally visual cues, all of which report on their actions. Thus, information is available from four distinct sensory channels—auditory, visual, tactile, and proprioceptive—contributing to a multisensory spatial coding of the event relative to the player's body (Kitagawa and Spence, 2006). Furthermore, with growing experience on the instrument, pianists gradually develop a multisensory cognition that, in some circumstances, even triggers involuntary perception-action mechanisms (Haueisen and Knösche, 2001).

Prior investigations made on a Yamaha Disklavier grand piano have offered partial insights into the aforementioned questions (Fontana *et al.*, 2017). Nevertheless, these studies have underscored the significance of further research in this domain. Disklavier pianos can play their keys autonomously, without requiring a pianist, thanks to servomechanisms controlled via the Musical Instrument Digital Interface (MIDI) protocol. This feature makes them exceptionally

[a)]Email: federico.fontana@uniud.it

valuable for exploring the audio-visual perception of tones in passive tasks. Specifically, in the mentioned experiment a group of pianists were required to localize a tone under three different conditions: (a) while the respective key was manually pressed by the pianist, (b) while the key was observed during its automatic movement, or (c) when they could only hear the produced tone. In all three cases, the tones were delivered binaurally through headphones, allowing for the random swapping of the auditory scene between the left and right channel across trials. While the results did not provide conclusive evidence to fully disentangle the roles of audition, vision, and proprioception, they suggest that visual and then somatosensory cues provide supplementary support to the auditory localization process. Indeed, their contribution ceased when the left and right binaural channels were reversed: in this case pianists, regardless of the sensory modalities enabled during conditions (a), (b), or (c), tended to indicate an approximate central position of the keyboard. This decline in localization performance suggests that, even when working in synergy as in condition (a), neither visual and proprioceptive feedback nor the memory of the multisensory event were robust enough to maintain the reversed sound source attached to the key producing the tone.

Based on this observation, our paper hypothesizes that piano sounds contain auditory cues supporting their perceptual localization. Our research aims to assess localization accuracy among musically experienced participants, and to identify the specific auditory cues, if any, involved in this process.

Testing the above hypothesis is not straightforward. Perhaps counter-intuitively, the acoustic field emitted by a piano is not guaranteed to contain lateralization cues. Certainly, such cues are absent during a tone's free evolution after the initial transient: in fact, the acoustic field results from the radiation of the soundboard, which is a large spatially distributed vibrating source. Vibrations of the soundboard caused by the mechanical wave originating in the string after hammer excitation occur in different regions, depending on the frequency of the corresponding oscillatory modes. This determines a continuous variation in the intensity of the tone's partial components around the pianist (Fletcher and Rossing, 1991, Sec. 12.5.2), hindering the localization of a freely evolving tone. In a study where piano tones were recorded using a microphone array and then played back to listeners through a loudspeaker array, it was found that cross-wiring the channels during playback did not significantly affect the perceived position of the piano relative to the listener in most configurations (Fontana *et al.*, 2013). This finding is surprising, given that inter-aural intensity differences are a primary cue for sound localization. One possible explanation lies in the physical characteristics of upright pianos: The soundboards of such instruments are typically rectangular with trimming rims, limiting vibrations to a roughly trapezoidal section across which the bridges run diagonally. An analysis of the corresponding vibrating modes (Fletcher and Rossing, 1991, Sec. 12.5.3) suggests that the trapezoidal shape of the active section of the soundboard may minimize inter-aural intensity differences during a tone's free evolution. Consequently, the primary source of lateralization cues in a piano tone likely resides in its initial transient.

A key study on piano transients by Askenfelt (1993) revealed that the *touch precursor*—the brief sound preceding the hammer's impact on the string, typically lasting 20–25 ms—contains a distinctive "thud" noise radiating from the keybed. This noise, contributing to the piano's characteristic timbre (Fletcher and Rossing, 1991, Sec. 12.9), originates from the key striking the keybed's wooden structure. Since this impact occurs at a specific location on the keybed, its sound could potentially trigger the auditory precedence effect (Blauert, 2013; Brown *et al.*, 2015; Carlile *et al.*, 1997; van der Heijden *et al.*, 2019). This effect enables listeners to almost instantaneously localize a sound source based on interaural time difference cues from the wavefront onset. Yet, its activation within the piano's acoustic near-field remains uncertain, since mechanical waves propagate faster through the keybed's rigid wood compared to airborne waves radiating from the impact point.

Little is known about the directional sound radiation of a piano, except for some grand models with their lid either closed or open (Brandner *et al.*, 2020; Fletcher and Rossing, 1991, Sec. 12.9.2). Especially in the higher frequencies, such waves are efficiently radiated not only by the soundboard but also by other parts of the instrument, making the diffusion of tone onsets a complex phenomenon (Tan *et al.*, 2018). Depending on the propagation speed of the former waves, the distribution in space of the resulting airborne wavefront onset radiating from the keybed may be too diffuse around the pianist to pinpoint a distinct impact location, thus challenging the activation of the auditory precedence effect. If, conversely, these transients indeed exist and can be isolated and decoded by a pianist, then recent measurements of near-field head-related transfer functions (HRTFs) offer valuable insights into the localization of nearby sound sources, accounting for azimuth, distance, and fundamental frequency (Li *et al.*, 2023; Parseihian *et al.*, 2014; Santarelli *et al.*, 1999; Shinn-Cunningham *et al.*, 2000; Yu *et al.*, 2018). Still, beyond the challenge of localizing sources during the free evolution of piano tones, their initial transients present a special case of near-field localization, given the covariance of azimuth and distance of a key location along the keyboard.

In addition to its relevance as a perceptual effect for pianists, tone localization holds relevance also for piano makers, especially those targeting models not intended for the concert hall. Upright and digital pianos serve as appealing alternatives to grand instruments, offering a more compact, cheaper and versatile choice. Some high-end digital pianos utilize multi-channel banks of meticulously recorded samples, creating a realistic soundfield through sophisticated loudspeaker systems (Koseki *et al.*, 2003). However, constrained by budget or space, many practitioners opt for compact digital pianos or software plug-ins, often connected to affordable stereo loudspeaker sets (e.g., those onboard some digital piano models), or to headphones (Yamana, 1986). Incorporating a realistic tone localization model in such cases could enhance the synthetic auditory scene, replicating essential spatial cues found in acoustic pianos.

The chosen Disklavier upright piano provides the required technology for our planned tests. Furthermore, our investigation potentially extends to grand pianos, suggesting that detecting auditory lateralization cues in upright piano tones implies their likely presence in grand pianos as well.

II. EXPERIMENT

An experiment was conducted to assess pianists' accuracy in localizing an upright piano tone using auditory feedback alone.

A. Setup

The experiment took place in a small, soundproofed recording studio (volume 42 m^3, noise rating NR 15 dB, RT60 $\approx$ 0.3–0.4 s). A Yamaha DU1A Disklavier upright piano was positioned against a wall in central position, with participants seated comfortably in front of it using a piano stool. Their head centers were approximately 82 cm from the keyboard front and 70 cm from the upright panel. The piano, covered with an acoustically transparent black cloth to obscure the keys, was MIDI-controlled by an Apple MacBook Pro laptop through an RME Fireface UCX audio interface. The Disklavier's servomechanism, disconnecting the strings from the hammers, was managed by an Arduino UNO microcontroller connected to the laptop via USB.

Two Genelec 8040 A loudspeakers, positioned at either sides of the instrument at 78 cm height (measured at the bottom), were angled at 47° to point toward the stool (see Fig. 1); the speakers were connected to two outputs of the audio interface. Both the stool and speakers were fixed in position throughout the experiment. Additionally, a Novation Nocturn MIDI controller was configured to offer three controls: a "Next" button, a "Repeat" button, and an endless knob, whose function is detailed below.

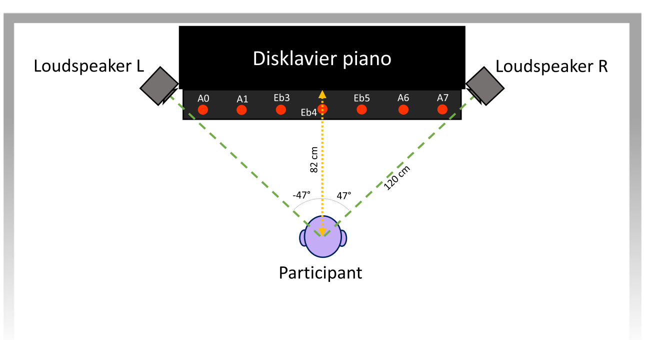

FIG. 1. (Color online) Experimental setup (top view). The participant's head (corresponding to VBAP center point) stood approximately at 82 cm from the piano's upright panel (71 cm from forehead + 11 cm to ears/head's center).

A software developed in PURE DATA (IEM, 2024) managed the entire experimental procedure, namely, by sending MIDI note data to operate the respective piano keys, switching the servomechanism to detach/connect the strings, generating synthetic stimuli as explained in Sec. II B, and finally processing and recording participants' responses.

B. Design and stimuli

The experimental design involved two crossed factors. In order to isolate the effect of key impact noise from a complete tone sound, *Tone* was defined as a categorical factor with two conditions: *full tone* and *transient tone*, representing the complete tone and key impact noise along with residual key mechanical noise, respectively. The factor *Key* featured seven levels corresponding to the notes A0, A1, Eb3, Eb4, Eb5, A6, and A7, positioned symmetrically around the center point Eb4 (Fig. 1). Notably, the respective keys were not equidistantly spaced, with distances from Eb4 of 16.5 cm (Eb3 and Eb5), 42 cm (A1 and A6), and 58 cm (A0 and A7). Their relative azimuth with respect to the vertical plane were –35.3, –26.7, –11.4, 0, 11.4, 26.7, and 35.3°. The corresponding MIDI key numbers are 21, 33, 51, 63, 75, 93, 105, respectively.

Full tone stimuli were generated by sending MIDI notes at *mezzoforte* intensity (MIDI velocity = 63) to the Disklavier in its standard configuration, i.e., with strings struck by the hammers. Transient tone stimuli followed the same process but involved decoupling the strings from the keyboard, thus producing key impact noise without strings vibrations.

Control sounds were synthesized for each stimulus:

- For full tone stimuli, exponentially decaying sine waves were generated, whose frequency and amplitude matched respectively the note's fundamental frequency and initial loudness. The frequency of the control sine wave for A0 (27 Hz) was doubled to 54 Hz due, on the one hand, to limited response of the loudspeakers (rated down to 45 Hz) and, on the other hand, to reduced hearing sensitivity at such low pitch. As a result, the control signals for A0 and A1 had the same pitch.
- For transient tone stimuli, a sound was created by combining three short recordings of hits against a wooden surface, each approximating a single impulsive event which forms the respective transient tone (Fig. 4 displays the characteristic intensity profile for each transient).

The control sounds were spatialized using *vector-base amplitude panning* (VBAP) (Pulkki, 1997) and reproduced through the loudspeakers. VBAP was chosen as a response method capable of preserving the auditory-only nature of the task. Alternative methods, such as hand pointing, might also involve vision and/or proprioception during localization. The VBAP focal point was set to the participants' head center, allowing lateralization of control sounds with a desired angular position. As detailed in Sec. II A, the distance from the upright panel to a listener's forehead was about 71 cm, with an additional 11 cm to reach their ears and head centers, totaling about 82 cm.

C. Participants

A total of N = 23 participants (11 males, 12 females) aged 19 to 46 (M = 28.5; SD = 7.5) were recruited from students and afferents of Zurich University of the Arts (ZHdK). Except for one participant, all had significant musical experience, with 15 having many years of piano training (M = 8.9, SD = 7.1). Two participants, with piano experience ranging from 2 to 25 years, reported absolute pitch. Participants signed an informed consent and received a 20 CHF voucher as compensation. All data were handled anonymously and securely. The study was conducted in accordance with the Declaration of Helsinki, and an informed consent was obtained from all subjects involved in the study.

D. Hearing balance test

Prior to the main test, participants underwent a screening for hearing balance between their left and right ears. This test collected participant-specific binaural offsets, and enabled us to take into account pitch-related variations in these offsets. Additionally, this phase familiarized participants with the experimental procedure used in the subsequent localization test.

The hearing balance test utilized the six control tones prepared for the main experiment. In this case, however, participants wore audiometry headphones (Sennheiser HDA 300), which allowed feeding their left and right ears separately. Before wearing headphones, participants removed glasses, earrings, etc. In each trial, the level of one channel was kept constant, while that of the other channel, initially set to zero, was adjusted by participants until localizing the stereophonic sound at the center. During adjustment, the control tone could be played as many times as necessary. Two measurements were taken for each ear, resulting in 24 trials (6 [pitch] $\times$ 2 [left] $\times$ 2 [right]) in random order. Sessions lasted about 10–15 min.

Offsets measured on both sides were expressed relative to the right channel and averaged for each participant and pitch. The grand mean offset across all participants and pitches was –0.15 dB (sd = 1.59) and did not significantly differ from 0 (t-test: p = 0.6497, t = –0.46046, df = 22). Pitch-specific mean offsets were within ± 1 dB except for Eb3 (m = –1.228, $p < 2.2 \times 10^{-16}$). Consequently, there was no systematic effect of pitch on the offsets.

Participant-specific hearing balance data were anonymized and recorded for later use in the analysis of localization performance.

E. Localization test

Participants then proceeded to the main localization experiment. The 7 [keys] $\times$ 2 [tones type] = 14 factor combinations were each measured four times, resulting in 56 trials and a session duration of 30–35 min. Presentation order was randomized within repetition rounds. Before each trial, the Disklavier's servomechanism was configured according to the tone type. Next, the piano played the stimulus, followed by the respective control stimulus through the loudspeakers. The task was to adjust the angular position of the control stimulus by turning the endless knob on the MIDI controller, until it matched that of the stimulus. While adjusting the position, participants could repeat the stimulus/control sequence at will by pressing the "Repeat" button on the MIDI controller. Then, they proceeded to the next trial by pressing the "Next" button.

III. RESULTS

The mean difference between responses and key position—here referred to as localization error—was modeled by Bayesian inference in R, using the brms package (Bürkner, 2017; Kruschke, 2014; R Core Team, 2021). Mean localization error μ was modeled by the azimuth angle of key location (in degrees), tone type, and their interaction. A distributional model estimated variable residual standard deviations σ for response distributions at each azimuth angle. Using a notation similar to that of generalized linear models, the model was specified as follows:

$$\begin{aligned} &\text{Error } \mu|\text{cens}(\text{censored}) \sim \text{Azimuth} * \text{Tone} \\ &\qquad\qquad\qquad\qquad + (1 + \text{Azimuth}|\text{sID}), \\ &\sigma \sim \text{Azimuth}, \end{aligned} \tag{1}$$

where responses at boundaries (–47°/47°) were treated as left or right censored data points and modeled in the term "cens(censored)." Individual intercepts and effects of "Azimuth" were specified as random effects for each participant (sID). The model was fit using a Gaussian distribution family. Stimuli, response dataset and analysis code are made available in an online repository (Fontana *et al.*, 2024).

A. Localization accuracy

The mean signed localization error served as a measure of localization accuracy. In Fig. 2 and Table I we present the estimated mean signed localization errors and their 95% Credible Intervals (CIs).[1] The data indicate that both full tones and transients were generally localized near the respective key position, yet some systematic deviations were observed.

Transient tones were localized nearly as well as full tones. Bayesian hypothesis tests revealed that errors were credibly larger for transient tones at the middle key Eb4 and the highest key A7: they were perceived 4.7° to the right and 4.3° to the left relative to full tones, respectively.

Mean responses deviated from the actual key locations most notably at the two lowest and highest keys (A0, A1 and A6, A7), localized up to 8.96° towards the center key position (see Table I). This lateral compression was unlikely to result from a ceiling effect due to the limited response range. Generally, less than 2% of responses were at either limit, and they were modeled as censored data points. The

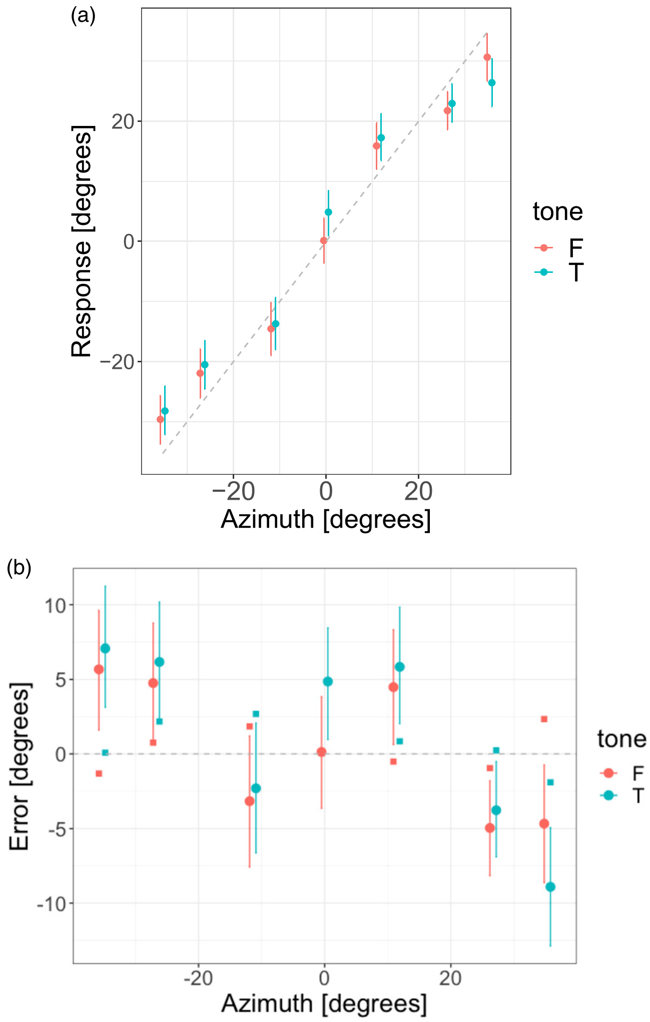

FIG. 2. (Color online) Localization responses (top) and errors (bottom) as a function of azimuth angle and tone type. F = full tones, T = transients; dotted line: ideal fit where localization = key location. Error bars = 95% Credible Intervals of estimated means. Square symbols in the bottom panel illustrate localization errors corrected according to Frank (2014), as discussed in Sec. IV A.

TABLE I. Estimated mean localization errors μ, with 95% credible intervals.

Key	Azimuth	Tone	Error	l-95% CI	u-95% CI
A0	−35.3°	F	5.54	1.36	9.49
A0	−35.3°	T	7.17	3.01	11.05
A1	−26.7°	F	4.73	0.80	8.69
A1	−26.7°	T	6.27	2.14	10.02
Eb3	−11.4°	F	−3.19	−7.72	1.23
Eb3	−11.4°	T	−2.38	−6.70	2.27
Eb4	0°	F	0.13	−3.68	3.96
Eb4	0°	T	4.85	1.08	8.60
Eb5	11.4°	F	4.50	0.60	8.38
Eb5	11.4°	T	5.94	2.12	9.71
A6	26.7°	F	−4.93	−8.28	−1.78
A6	26.7°	T	−3.69	−6.87	−0.47
A7	35.3°	F	−4.62	−8.49	−0.66
A7	35.3°	T	−8.96	−12.78	−5.04

only exception was the full tone A0 which had about 10% of the responses at the left boundary, but this would not explain the effect at A1, A6, and A7, or the transient tone at A0.

The three middle keys (Eb3, Eb4, and Eb5) were localized nearest to the actual key position. While the full tones at Eb4 were on average localized with a negligible mean error of –0.13°, the locations of both full and transient tones at Eb3 and Eb5 were slightly overestimated; however, the CIs for Eb3 do not support credibly non-zero effects. Additionally, the lower CI for Eb5 full tone is near zero.

Left-right hearing balance, assessed in the preliminary test, did not yield a systematic effect on localization results. However, a weakly positive yet significant correlation was observed between the unsigned mean offsets for a given key and respective unsigned localization errors (Pearson's $r = 0.07$, $p < 0.05$), potentially indicating individual differences in localization ability. Piano experience in years did not show a significant correlation with localization accuracy for either full tones or transients. Absolute pitch had no discernible systematic effect either: one absolute pitch possessor performed better with full tones than transients, while the other performed worse.

B. Precision

To assess differences in localization precision, standard deviations (σ) of the distributions of signed errors were estimated for each azimuth in the statistical model given in Eq. (1). Results are provided in Table II. The estimated standard deviation exhibits a slight dependence on azimuth. The smallest estimate ($\sigma = 7.50$), i.e., the best measured precision, is observed at azimuth 0°. The estimate increases towards the lateral keys, reaching $\sigma = 9.7$ and $\sigma = 9.39$ for azimuths –35.3° and 35.3°, respectively. Hypothesis tests between the lowest key A0 and each other key reveal that the estimates are credibly smaller for central positions within the azimuth range [–11.4°, 26.7°]. By contrast, the estimates for the more lateral positions A1 (–26.7°) and A7 (35.3°) do not differ credibly from A0 (−35.3°).

IV. DISCUSSION

The results raise several points of discussion concerning the nature, variety and use of the lateralization cues existing in both the stimuli and respective control signals.

TABLE II. Estimated residual standard deviations σ of the response distributions for each azimuth angle, with 95% credible intervals.

Key	Azimuth	σ Est.	l-95% CI	u-95% CI
A0	−35.3°	9.70	8.66	10.88
A1	−26.7°	8.96	8.03	9.99
Eb3	−11.4°	7.60[a]	6.84	8.48
Eb4	0°	7.50[a]	6.76	8.34
Eb5	11.4°	7.98[a]	7.18	8.90
A6	26.7°	7.99[a]	7.20	8.88
A7	35.3°	9.39	8.43	10.51

[a]Credible difference from a hypothesis test against azimuth –35.3°.

https://doi.org/10.1121/10.0026484

FIG. 3. (Color online) Recording setup with dummy head binaural microphone.

A. Effect of source reproduction

The compression of perception towards the center of the keyboard for the most lateral keys (A0, A1, A6, and A7) contrasts with the accurate localization of the full tone Eb4 at azimuth 0°. This compression could not be explained through a ceiling effect. Interestingly, the keys Eb3 (–11.4°) and Eb5 (11.4°) were in turn slightly more lateralized. Both discrepancies might be attributed to deviations between real and VBAP-panned sources, as measured and predicted in previous studies (Frank, 2014). The square notation in Fig. 2 (bottom) demonstrates how correcting the measured localisation errors according to Frank explains the systematic errors almost entirely. The co-variation of distance and azimuth adds complexity to the perception of nearby sources: previous studies reporting over- and underestimation of distances for specific ranges ($d < 0.75$ m and $0.75 < d < 1$ m) (Brungart, 1999; Parseihian *et al.*, 2014) highlight this

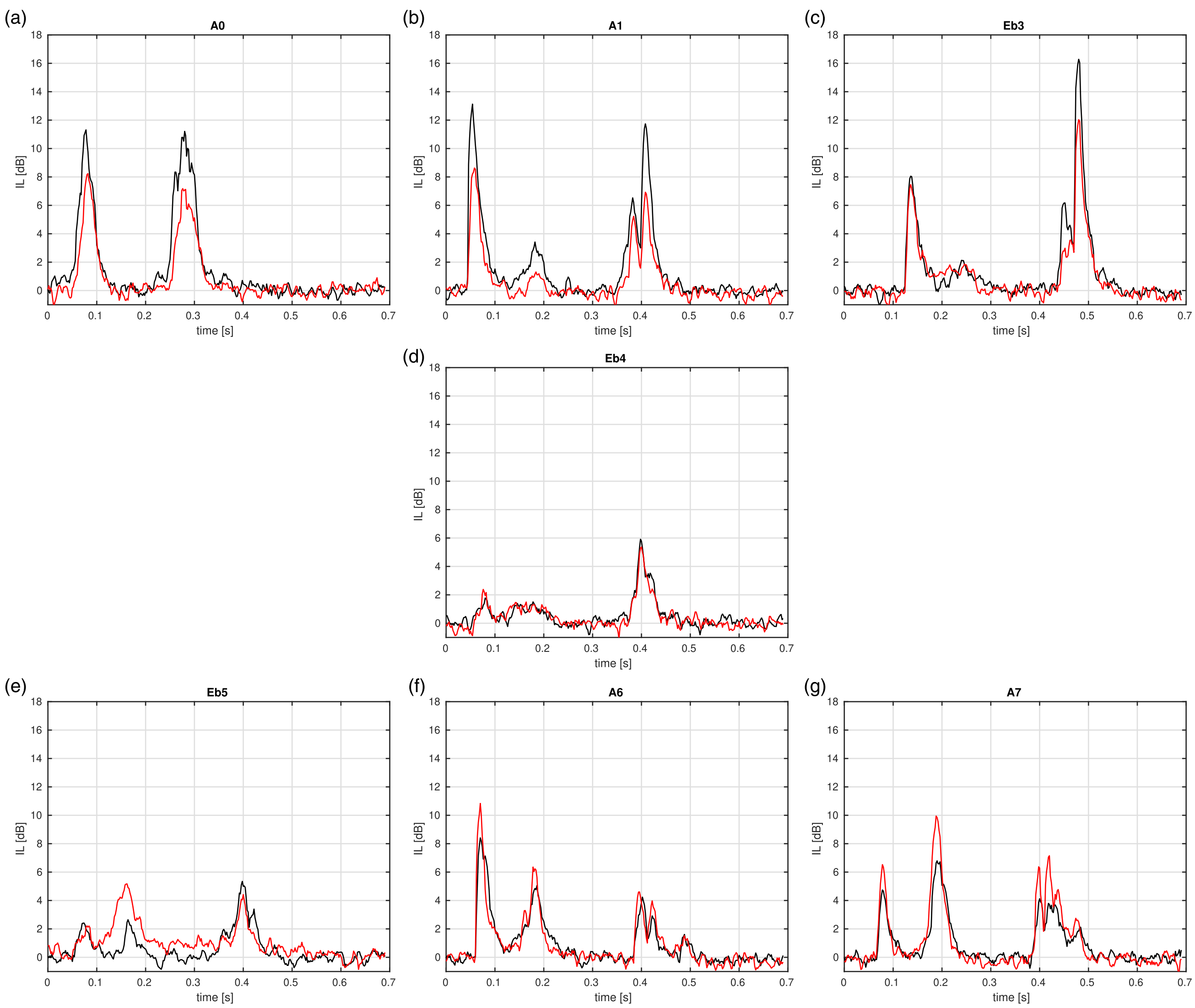

FIG. 4. (Color online) IL after noise floor subtraction of transients in notes A0, A1, Eb3 [(a)–(b)], Eb4 (d), Eb5, A6, A7 [(e)–(g)]. Left binaural sound in black colors; right binaural sound in red color.

challenge. The simultaneous influence of distance localization and lateralization may contribute to the observed overshoot and undershoot patterns in our results.

In terms of precision, our results align with previous findings on nearby brief Gaussian broadband noise bursts, which reported mean absolute azimuth error of approximately 6.1° to 6.5° (sd 7.4 to 7.8) at frontal positions (Parseihian *et al.*, 2014). By comparison, our study yielded mean absolute azimuth errors of 7.16° (sd 7.63) for full tones and 7.81° (sd 6.86) for transient tones. These precisions are somewhat worse than reported minimum audible angles (MAA) (Mills, 1958). For instance, Meng *et al.* (2021) found MAA for VBAP-produced virtual broadband white noise bursts ranging from 1.1° to 3.1° for azimuths of 0° and 90°, respectively. However, due to differences in experimental designs, pointing methods, and the MAA measurement approach (adaptive algorithm finding the 79.4% point on the psychometric function), a strict comparison is not feasible.

While piano sound radiates from the instrument's rectilinear body, VBAP sources lie on a circle intersecting the loudspeakers and centered on the listener's head. This geometric difference introduces a distance discrepancy between the physical and VBAP sources, potentially influencing the apparent source width (Griesinger, 1997). In our setup, the VBAP domain has a radius of 120 cm, and the corresponding geometric difference ($120 - 82 = 38$ cm) occurs when both the instrument and VBAP sources are directly in front of the listener, i.e., when Eb4 was played and the related stimulus was correctly localized around the corresponding key. Assuming the apparent width of a source radiating from the instrument is $2w$ cm, moving this source farther from 82 to 120 cm reduces its apparent width from $2\arctan(w/82)$ to $2\arctan(w/120)$ degrees. For instance, supposing that the apparent width of the perceived physical source covered the largest range around Eb4 beyond which localization confusion would start (corresponding to an inter-key distance amounting to 16.5 cm), then moving this source farther would decrease its apparent width from about 22° to about 15°. Both of these widths are approximately twice as wide as the average precision our participants showed in the experiment, indicating that distance bias did not affect localization when Eb4 was played. Even more so, apparent source width differences should progressively diminish towards the keyboard's ends; hence, they should impact subjective precision proportionally. However, this was not observed, as discussed in Sec. III B, suggesting low or no influence of distance differences between physical and VBAP sounds on localization.

B. Effect of interaural differences

The localization of transient tones by our participants suggests a reliance on interaural level difference (ILD) and interaural time difference (ITD) cues for lateralization. To investigate this further, we acquired binaural recordings of the transient tones at $F_S = 192$ kHz sampling frequency using a Neumann KU100 dummy head binaural microphone. The microphone was positioned at the average head location of participants (see Fig. 3).

The resulting digital samples were processed by applying a third-order Butterworth bandpass filter in the range [2000, 20 000 Hz], which is standard in ILD analysis. This was done through the MATLAB^{TM} function `butter`. Intensity levels (IL) of the filtered signals were computed every 2 ms across 10 ms windows, i.e., with an overlapping factor equal to 5. The residual noise floor IL was determined for each channel during a period of piano inactivity, and then subtracted from the respective IL.

Figure 4 displays the IL of all transient tones for each note after noise floor subtraction over 700 ms. It is worth noting that each transient comprises three impulsive events: the initial two occur when the hammer strikes the keybed, while the third happens when it rebounds to the rest position. Observable differences between the left and right levels are evident in the plots, with variations reaching

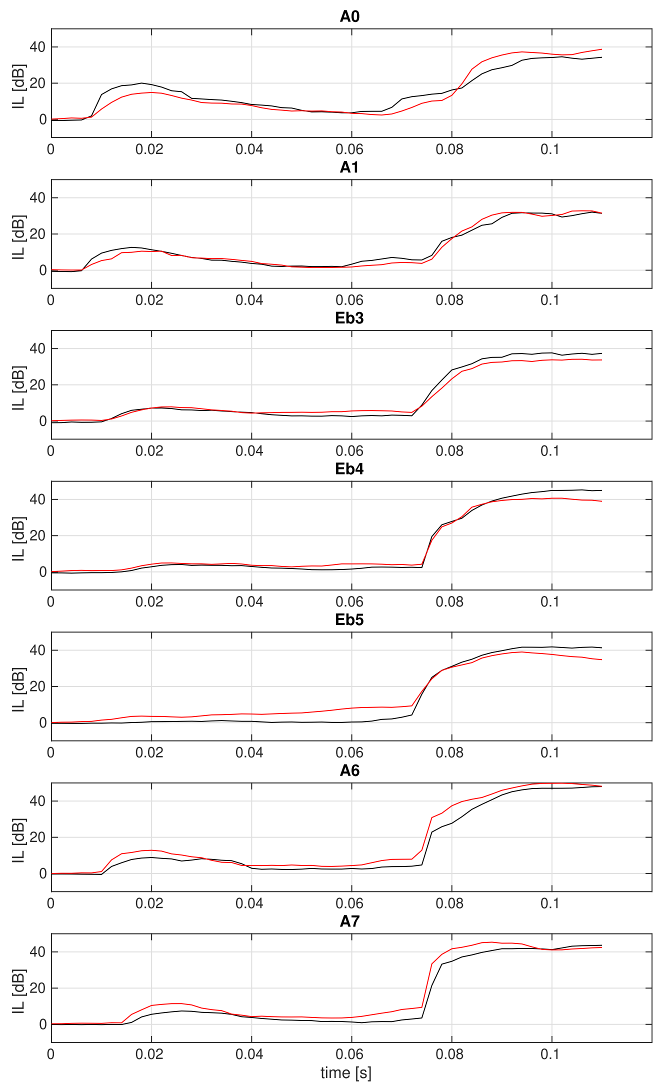

FIG. 5. (Color online) IL after noise floor subtraction of full tones, for the first 100 ms across all notes. Left binaural signal represented in black color; right binaural signal in red.

approximately 4–5 dB at the most prominent ridges of the transients. These differences gradually diminish as the stimulus moves to the center.

Through testing with the Neumann dummy head, Watanabe *et al*. (2007) established relationships between ILD and source directions. For our measured IL, this corresponds to angles ranging from 10° to 35° across different frequency bands, which align with our listeners' judgments. The same authors demonstrated subjective differences between the relations obtained with the dummy head and those resulting from tests with human listeners. We infer that our participants were, in principle, capable of localizing note positions through ILD cues in the transients. In particular, the slight rightward offset observed when the Eb4 tone was played is consistent with an ILD of approximately 1 dB, as evident in the corresponding plot in Fig. 4 at around 80 ms.

As predicted by Askenfelt, our transients contribute to the formation of touch precursors limited to the first few milliseconds before the string sounds are radiated. To verify this in our setup, full piano tones were sampled at $F_S = 192$ kHz with the Neumann dummy head and then filtered as described above for the transient tones. Figure 5 shows the IL of the initial 100 ms for each full tone. Precursors appear in each plot, centered around approximately 10 ms, suggesting that ILD cues are conveyed also by full tones before the piano radiates the louder string sounds. The signals corresponding to the precursors leading to Fig. 5 are depicted in Fig. 6, where spectral components outside the audible band (here conventionally set to [20, 20 000 Hz]) were preliminary filtered out.

A potential role of ITD cues in forming precedence effects has been suggested, with Pastore (2020) proposing that "…the precedence effect may be conceptualized as instances of observed onset dominance that are integrated across different time scales and levels of auditory processing to form an overall […] auditory percept that is dominated by the first arriving stimulus." In line with this hypothesis, we specifically examined the transient tone onsets for possible ITD cues.

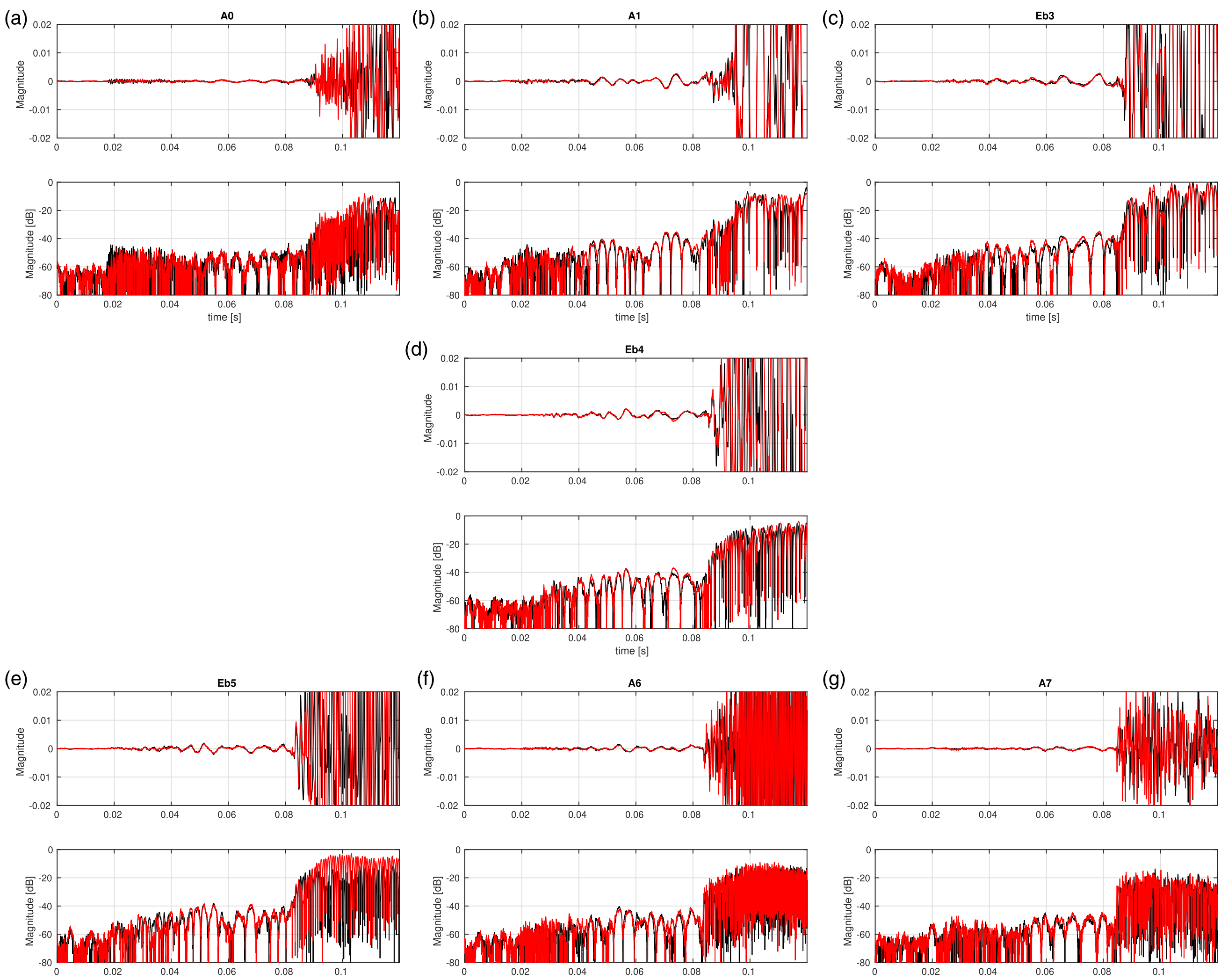

FIG. 6. (Color online) Full tone attacks after noise floor removal in notes A0, A1, Eb3 [(a)–(b)], Eb4 (d), Eb5, A6, A7 [(e)–(g)]. Upper subplot: magnitude; lower subplot: dB magnitude. Left binaural signal in black color; right binaural signal in red.

Multichannel recordings of the transients were obtained by using a seven-channel microphone array placed in at the participants' head average position, as shown in Fig. 7. Every microphone (iSEMcon EMM-7101-CSTB) was positioned 9 cm apart from the next one, resulting in a total array length of 55 cm. The recordings, sampled at $F_S = 192$ kHz, allowed capturing the soundfield at the microphone positions with an equivalent spatial resolution of $c/F_S \approx 1.7$ mm between subsequent pairs of pressure field values, assuming the speed of sound $c = 343$ m/s.

The multichannel signals were processed using zero-phase digital filtering, by applying a third-order Butterworth bandpass filter in the range [20, 2000] Hz through the MATLAB™ function `filtfilt`. Figure 8 shows the initial 10 ms of each signal, approximately corresponding to the onset of the respective transient tone, as evident in Fig. 5. Temporal trends emerge, revealing varying trajectories across the seven channels. These trends occasionally reveal inter-channel delays consistent with those produced by a wavefront approaching the listening points from the positions of the targeted piano keys. Recalling that the black and red lines denote the leftmost and rightmost microphones, respectively, with a distance corresponding to the array length, a few observations can be made. For example, in the mid plots of Eb4 (Fig. 8), between 4 and 6 ms, an ascending trajectory is detected by the rightmost microphone with a relative delay of about 0.1 ms. Similarly, in the top-left plots of A0 signals, a trend propagates with a relative delay of about 1 ms during the same temporal window.

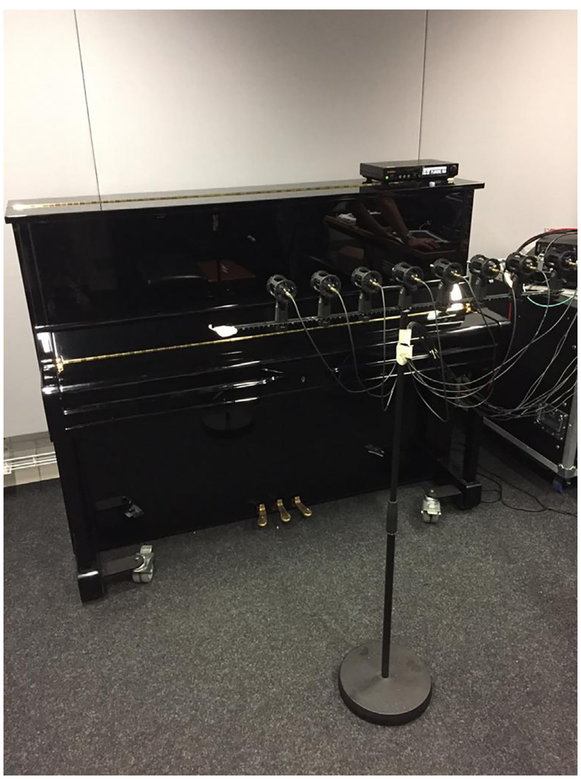

FIG. 7. (Color online) Recording setup with microphone array.

These delays in fact indicate radiating sources that, if assumed to be point-wise, would be situated at varying distances from the center of the keybed. Calculating such distances involves solving an Euclidean geometry problem using the measured delays. In Table III, for each note, we present the temporal window range where a trend was observed, the measured rightward movement delay, and the calculated key distance from the center of the keybed, along with the direction of the corresponding sound source. From this analysis, limited to the notes that were object of the tests, we conclude that the touch precursors in our piano convey ITD cues that, within the limits of their accuracy, support sound source localization.

C. Effect of dynamic localization

Since our participants could slightly orient their heads during each trial, dynamic localization likely played a role in their performance. The advantages of dynamic localization vs settings where the listener's head cannot be rotated are well-documented in the literature (Thurlow and Runge, 1967; Wallach, 1940). More recently, head movements in normal listeners performing a pointing task have been shown to be useful, although not strictly necessary for achieving greater sound source localization accuracy (Gulli *et al.*, 2022). However, in our setting we did not track head rotations, as they worked in synergy with dynamic lateral adjustment of the control sounds, making it difficult to disentangle their respective impacts. Moreover, significantly reduced lateral errors were measured for head rotations up to 32° azimuth (McAnally and Martin, 2014). Considering our stimuli, which may not have provided robust localization information, listeners might have engaged in a series of differential measurements, seeking an angle with null interaural differences, and completing the trial when they perceived sound lacking directional cues.

D. Effect of audio-visual imagery

The marginal improvement in localizing full tones compared to transients suggests the involvement of learned directional cues based on audio-visual imagery of key locations. To isolate the role of these higher-level cues and assess their salience, future studies could consider canceling lateralization cues found in the onsets of stimuli with non-directional noise. In this regard, the literature provides evidence that simply hearing a piano tone can trigger representations of the corresponding physical movement in pianists (Novembre and Keller, 2014). Specifically, Taylor and Witt (2015) showed that pianists (but not novices) develop coarse spatial representations when listening to ascending and descending scales. However, the insignificant correlation between years of piano experience and localization accuracy we reported in Sec. III A suggests that the impact of such imagery may be limited compared to direct auditory localization cues.

It would be compelling to measure localization performance using piano tones without key mechanical noise.

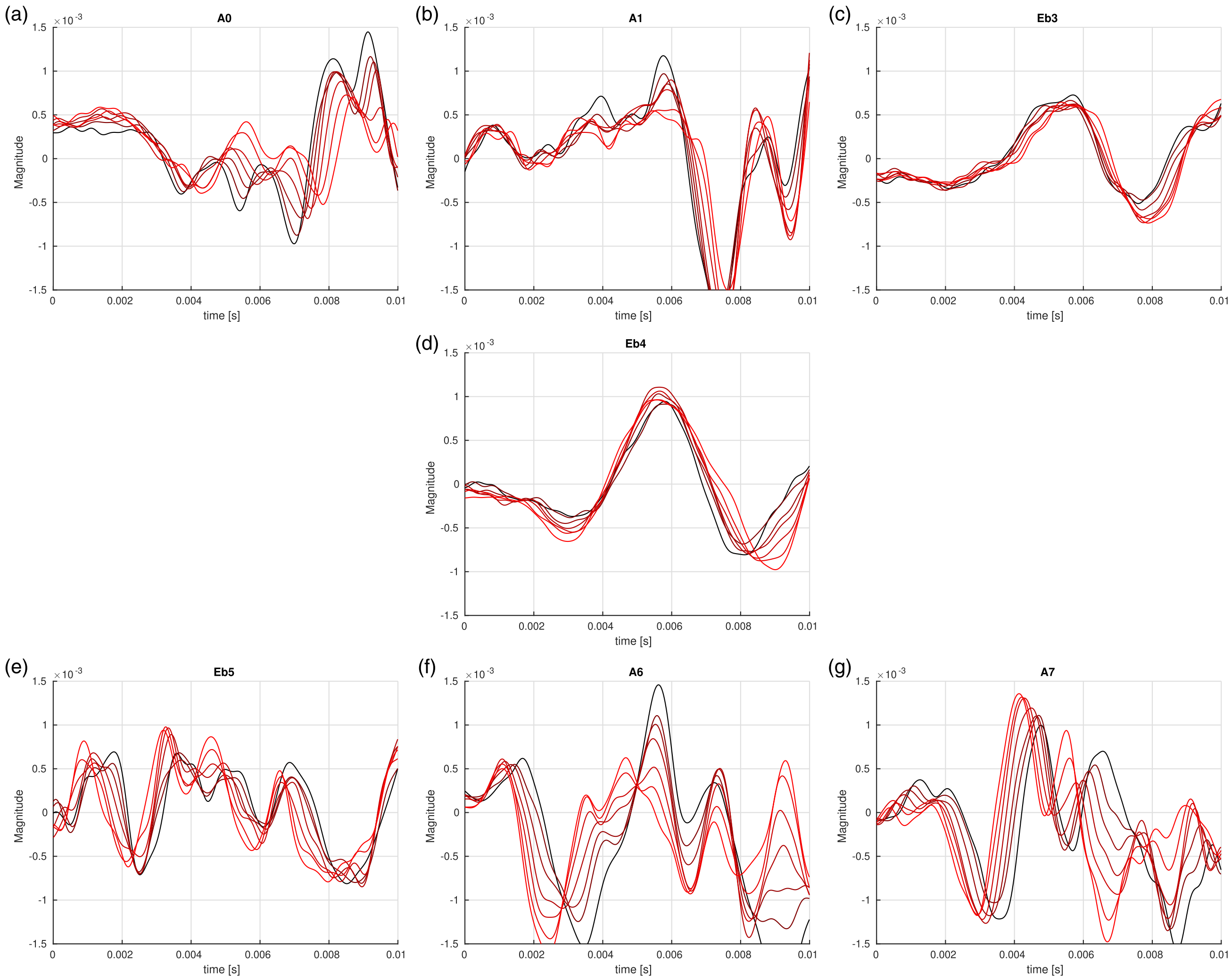

FIG. 8. (Color online) Multichannel recording of touch precursors in notes A0, A1, Eb3 [(a)–(c)], Eb4 (d), Eb5, A6, A7 [(e)–(g)], first 10 ms. Left to right microphone signals are represented with colors transitioning from black to red.

Unfortunately, limitations of our Disklavier setup—specifically inconsistent key response times (jitter) to MIDI note on/off messages—prevented accurate masking of key depression and especially release noises. This precluded a condition where the Disklavier produced no mechanical noise, even for low-intensity tones with minimal touch precursors (see Fig. 6). Furthermore, presenting auditory stimuli lacking informative precursors could have introduced unintended biases, providing once again just a partial answer about the role of memory in piano tone localization. Finally, our participant pool included only two individuals with absolute pitch, preventing a robust statistical analysis of its potential influence on localization.

TABLE III. Temporal window range, delay trend, resulting key distance from the keybed center, and resulting sound source direction for each note.

Key	Range (s)	Delay (ms)	Distance (m)	Direction (deg.)
A0	[0.004 0.006]	0.9	0.61	36.6°
A1	[0.006 0.008]	0.5	0.29	19.4°
Eb3	[0.006 0.008]	0.4	0.24	16.3°
Eb4	[0.006 0.008]	0.1	0.06	4.2°
Eb5	[0.002 0.004]	−0.3	−0.17	−11.7°
A6	[0.004 0.006]	−0.6	−0.33	−21.9°
A7	[0.004 0.006]	−0.7	−0.40	−26.0°

E. Other effects

Loudness and spectral differences among transient tones, as well as reverberation (if subtle) associated with their radiation toward the listener, might be acknowledged as potential non-directional cues, enriching the understanding of factors influencing the localization process. While such differences may decrease the correlation between intensity and spectral cues with sound source positions, they can characterize individual transients, allowing listeners to attach acoustic labels to each repetition across trials. These labels may contribute to some of the systematic effects on localization recorded among our participants. However, any such effects were potentially mitigated by the matching distance of test and control stimuli.

V. CONCLUSIONS

This study demonstrates that individuals with substantial musical expertise, including pianists, effectively leverage lateralization cues present in tone precursors to localize notes across a significant portion of the piano keyboard, as observed with our upright piano. Isolating these precursors revealed their central role in guiding the localization process through interaural cues. Their effectiveness even when embedded within complete tones suggests potential involvement of the precedence effect, aligning with Askenfelt's hypothesis (Askenfelt, 1993). However, further investigation of instrument sound radiation, possibly using specialized technologies (e.g., an acoustic camera), is needed to fully confirm this conclusion.

Especially in experienced musicians, full tones may evoke visuo-spatial imagery of the piano keyboard through the note's fundamental frequency. Pianists, in particular, might benefit from this association, linking a note to its corresponding key position. Our future research will explore the cognitive role of harmonic content in the piano tone localization process in greater detail.

ACKNOWLEDGMENTS

The authors acknowledge Federico Avanzini and Roberto Barumerli for their suggestions and contribution to the analysis of full and transient tones in Sec. IV B.

AUTHOR DECLARATIONS

Conflict of interest

The authors declare they have no conflicts of interest or competing interests to disclose.

Ethics approval

Informed consent was obtained from all subjects involved in the study. The study was conducted entirely in Switzerland, in accordance with the Declaration of Helsinki. However, it falls outside the scope of the Swiss Human Research Act, and therefore approval from the National Ethics Committee was not required for its implementation.

DATA AVAILABILITY

Stimuli, response dataset and analysis code are made available in Fontana *et al*. (2024).

[1]The Bayesian 95% Credible Interval gives the interval in which the parameter falls with the specified certainty.

Askenfelt, A. (**1993**). "Observations on the transient components of the piano tone," STL-QPSR **34**(4), 15–22.

Blauert, J. (**2013**). *The Technology of Binaural Listening* (Springer, Berlin).

Brandner, M., Meyer-Kahlen, N., and Frank, M. (**2020**). "Directivity pattern measurement of a grand piano for augmented acoustic reality," in *Fortschritte Der Akustik (DAGA)* [*Advances in Acoustics (DAGA)*], Hanover, Germany, pp. 138–141.

Brown, A. D., Stecker, G. C., and Tollin, D. J. (**2015**). "The precedence effect in sound localization," J. Assoc. Res. Otolaryngol. **16**(1), 1–28.

Brungart, D. S. (**1999**). "Auditory localization of nearby sources. III. Stimulus effects," J. Acoust. Soc. Am. **106**, 3589–3602.

Bürkner, P.-C. (**2017**). "brms: An R package for Bayesian multilevel models using Stan," J. Stat. Softw. Art. **80**(1), 1–28.

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**(1), 179–196.

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer-Verlag, New York).

Fontana, F., Järveläinen, H., Papetti, S., and De Pra, Y. (**2024**). "Piano localization dataset," https://doi.org/10.5281/zenodo.11654054 (Last viewed June 27, 2024).

Fontana, F., De Pra, Y., and Amendola, A. (**2013**). "Sensitivity to loudspeaker permutations during an eight-channel array reproduction of piano notes," in *Proceedings of SMAC/SMC 2013*, Stockholm, Sweden.

Fontana, F., Scappin, D., Avanzini, F., Bernardi, M., Bianco, D., and Klauer, G. (**2017**). "Auditory, visual and somatosensory localization of piano tones: A preliminary study," in *Proceedings of the 14th Sound and Music Computing Conference*, pp. 254–260, http://smc2017.aalto.fi/media/materials/proceedings/SMC17_p254.pdf (Last viewed June 27, 2024).

Frank, M. (**2014**). "Localization using different amplitude-panning methods in the frontal horizontal plane," in *Proceedings of EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany.

Griesinger, D. (**1997**). "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," Acustica **83**, 721–731.

Gulli, A., Fontana, F., Orzan, E., Aruffo, A., and Muzzi, E. (**2022**). "Spontaneous head movements support accurate horizontal auditory localization in a virtual visual environment," PLoS One **17**(12), e0278705.

Haueisen, J., and Knösche, T. R. (**2001**). "Involuntary motor activity in pianists evoked by music perception," J. Cogn. Neurosci. **13**(6), 786–792.

IEM, "Pure data," http://puredata.info/ (2024) (Last viewed June 27, 2024).

Kitagawa, N., and Spence, C. (**2006**). "Audiotactile multisensory interactions in human information processing," Jpn. Psychol. Res. **48**(3), 158–173.

Koseki, S., Mantani, R., Sugiyama, N., and Tamaki, T. (**2003**). "Method for making electronic tones close to acoustic tones, recording system and tone generating system" EU Patent no. EP1357538 A3.

Kruschke, J. (**2014**). *Doing Bayesian Data analysis—A Tutorial with R, JAGS, and Stan*, 2nd ed. (Academic Press, New York).

Li, Y., Giusto, F. D., Van Ophem, S., Preihs, S., Deckers, E., and Peissig, J. (**2023**). "Cross-validation of acquisition methods for near-field Head-Related Transfer Functions with a high distance resolution," Proc. Mtgs. Acoust. **49**, 032001.

McAnally, K. I., and Martin, R. L. (**2014**). "Sound localization with head movement: Implications for 3-d audio displays," Front. Neurosci. **8**(8), 210–210.

Meng, R., Xiang, J., Sang, J., Zheng, C., Li, X., Bleeck, S., Cai, J., and Wang, J. (**2021**). "Investigation of an MAA test with virtual sound synthesis," Front. Psychol. **12**, 656052.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**(4), 237–246.

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.

Parseihian, G., Jouffrais, C., and Katz, B. F. (**2014**). "Reaching nearby sources: Comparison between real and virtual sound and visual targets," Front. Neurosci. **8**, 269.

Pastore, M. T. T. (**2020**). "The precedence effect as onset dominance observed at multiple levels of processing? A modeling study," in *Proceedings of Forum Acusticum 2020*, Lyon, France, pp. 1195–1202.

Pulkki, V. (**1997**). "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc. **45**(6), 456–466.

R Core Team (**2021**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), https://www.R-project.org/ (Last viewed June 27, 2024).

Santarelli, S., Kopco, N., Shinn-Cunningham, B. G., and Brungart, D. (**1999**). "Near-field localization in echoic rooms," J. Acoust. Soc. Am. **105**(2), 1024.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**(3), 1627–1636.

Tan, J. J., Chaigne, A., and Acri, A. (**2018**). "Operational transfer path analysis of a piano," Appl. Acoust. **140**, 39–47.
Taylor, J. E. T., and Witt, J. K. (**2015**). "Listening to music primes space: Pianists, but not novices, simulate heard actions," Psychol. Res. **79**(2), 175–182.
Thurlow, W. R., and Runge, P. S. (**1967**). "Effect of induced head movements on localization of direction of sounds," J. Acoust. Soc. Am. **42**, 480–488.
van der Heijden, K., Rauschecker, J. P., de Gelder, B., and Formisano, E. (**2019**). "Cortical mechanisms of spatial hearing," Nat. Rev. Neurosci. **20**(10), 609–623.
Wallach, H. (**1940**). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**(4), 339–368.
Watanabe, K., Ozawa, K., Iwaya, Y., Suzuki, Y., and Aso, K. (**2007**). "Estimation of interaural level difference based on anthropometry and its effect on sound localization," J. Acoust. Soc. Am. **122**, 2832–2841.
Yamana, K. (**1986**). "Electronic musical instrument having a pan-pot function" U.S. patent 4,577,540 A.
Yu, G., Wu, R., Liu, Y., and Xie, B. (**2018**). "Near-field head-related transfer-function measurement and database of human subjects," J. Acoust. Soc. Am. **143**, EL194–EL198.