A Pseudo-Haptic Method Using Auditory Feedback: The Role of Delay, Frequency, and Loudness of Auditory Feedback in Response to a User's Button Click in Causing a Sensation of Heaviness

SEITARO KANEKO[1,2], (Student Member, IEEE), TAKUMI YOKOSAKA[3],
HIROYUKI KAJIMOTO[1], (Member, IEEE), AND TAKAHIRO KAWABE[3]

[1]Department of Informatics, The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan
[2]Japan Society for the Promotion of Science, Chiyoda, Tokyo 102-0083, Japan
[3]NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Kanagawa 243-0198, Japan

Corresponding authors: Seitaro Kaneko (kaneko@kaji-lab.jp) and Takumi Yokosaka (yokosaka.takumi@gmail.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the NTT Communication Science Laboratories Ethics Committee Approval No. R02-009, and performed in line with the Helsinki Declaration.

Seitaro Kaneko and Takumi Yokosaka contributed equally to this work.

ABSTRACT One of the important challenges in the field of haptic engineering is the development of methods to present haptic information on commonly-used information devices equipped only with screens and speakers. Earlier research has proposed the so-called ''pseudo-haptic'' method that can give users haptic impressions by modulating the visual feedback provided in response to user inputs. In this study, to extend the applicability of the pseudo-haptic method to devices without a screen, we propose a novel method for varying the heaviness sensation experienced by users by modulating auditory (as opposed to visual) feedback provided in response to user inputs. In this method, we manipulated the delay, frequency, and loudness of auditory feedback (a pure tone) given in response to the user clicking a button. Through a series of psychophysical experiments, we found that participants tended to report a stronger heaviness sensation when a pure tone was presented with a longer delay, lower frequency, and/or greater loudness. By manipulating the onset and offset timing of a pure tone, we also demonstrated that the delay of the offset of the pure tone, rather than that of its onset, was critical to the heaviness sensation. Our sound-based pseudo-haptic method can be implemented in information devices that can present auditory information, regardless of whether they can also present visual information or not.

INDEX TERMS Auditory feedback, cross-modal interaction, delay, heaviness, pseudo-haptics.

I. INTRODUCTION

A. BACKGROUND

When touching and handling an object, humans perceive haptic properties such as heaviness, softness, and roughness based on the object's physical properties such as density, elasticity, and surface friction. Previous studies in the fields of human-computer interaction and virtual reality have focused on developing haptic displays that present these haptic properties virtually by modulating stimulus inputs like force and vibration without changing the actual physical properties of real objects. For example, without changing the weight of an actual object, it is possible to impart the sense of various levels of heaviness by modulating force feedback in the direction of gravitational force using grounded kinesthetic haptic devices [1], [2]. Also, the sensation of softness can be imparted using a device that changes the contact area between finger and object [3], and the sensation of roughness can be imparted using a device that generates electrovibrations while users operate a touchscreen [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma.

Unlike the haptic devices described above, since most information devices such as smartphones and PCs only have screens and speakers (and at best a very simple vibratory function), conventional methods for modulating force and friction cannot be implemented in them. This is of course why it is not easy to present information about haptic properties in commonly-used information devices.

In earlier research, a method was devised whereby users experienced the sensations associated with various haptic properties, without the need for elaborate haptic devices, by taking advantage of cross-modal interaction. The technique involved in producing such haptic sensations in the field of human-computer interaction is called pseudo-haptics. The pseudo-haptic technique induces haptic sensations by modulating the visual feedback given in response to user input [5], [6]. For example, reducing the speed of cursor movement relative to the speed of mouse movement induces an illusory sense of heaviness [7]. Presenting a deformed image of an object that can be indented by mouse clicks or other means intensifies the sense of hardness communicated to the user [8], [9].

Although the phenomenon of pseudo-haptics has attracted considerable attention, there are limitations to its use. The conventional pseudo-haptic technique requires a monitor or screen to present visual information, and thus, does not work on information devices without one. If this limitation were to be overcome, the method could be made available in devices that do not have a screen and/or a transducer, such as remote controllers, earphones, and bracelets. The motivation for our study was to implement a method of inducing a haptic impression using information other than visual and haptic information.

Here we propose a novel type of pseudo-haptic technique for varying a sensation of heaviness by modulating auditory feedback instead of visual feedback. This method is easy to implement since many information devices are equipped with speakers. In our daily lives, we often experience situations where auditory feedback is presented, for example, when we press a button on remote controllers for TVs, air conditioners, and lights. Hence, it is reasonable to assume that auditory feedback is familiar and intuitive to users. If the pseudo-haptic technique using auditory feedback modulation is found to be effective in presenting a sense of heaviness, it may be combined with another pseudo-haptic technique using visual feedback modulation to achieve richer and more intuitive presentations of haptic properties in various scenarios.

In this study, we propose three elements of auditory feedback that can contribute to the pseudo-haptic technique: delay, sound frequency, and sound loudness. Delay refers to a temporal interval between an action (e.g., a button press) and the presentation of auditory feedback. Conventional studies investigating the effect of visual feedback on perception reported that a longer delay applied to visual avatars moved by a user's hand motions increased the impression of illusory heaviness, resistance, or hardness [10]–[13]. A recent study

has shown that illusory heaviness can be induced just by a user's key press [14]. In that study, when the delay between a user's key press and a corresponding change in luminance on a display was longer, participants reported a more pronounced sense of heaviness. A similarity in the effects of feedback delay on perception has been reported not only for visual but also for auditory feedback. For example, when a delay is inserted between a button press and the corresponding feedback, the perception of the association between the button press and the auditory feedback is weakened [15]–[17] just as it is for the association between a button press and corresponding visual feedback [18]–[22]. Based on these findings, we hypothesized that the sense of heaviness could be increased by manipulating the delay between a user's key press and the auditory feedback provided in response to it

The second and third elements of auditory feedback that can contribute to the pseudo-haptic technique are related to the frequency and loudness of the sound presented in response to a user's action. Objects having different physical properties tend to produce different sounds. For example, larger and heavier objects tend to produce louder and lower frequency sounds when colliding [23]. Humans might use the sound phenomena created by heavier objects to estimate heaviness. Some results that support this hypothesis have been reported. For example, participants tended to rate lower frequency sounds as heavier [24], and the frequency of synthesized sounds affected participants' perception of a material [25]. The heaviness impression induced by lower frequency sound has been shown to occur even for an object that participants are actually holding in their hands [26]. Based on these findings, we hypothesized that users would experience a greater sense of heaviness when a tone with lower pitch and/or greater loudness was generated in response to their action.

In what follows, we will first describe the relationship between our proposed method and earlier related studies on pseudo-haptics methods and auditory-haptic association. Then, we will describe our method and report on the psychophysical experiments that were conducted to evaluate its performance. Finally, we will discuss its limitations and future issues.

B. RELATED WORK
1) MULTISENSORY INTEGRATION IN A DIGITAL SETTING
Users can achieve rich sensory experiences by integrating information from multiple senses. When users interact with an external object with motor action, the action often causes a set of changes in the object. For example, the action may cause the object's movement, deformation, sound, and vibration. To perceptually comprehend what sort of the change in the object is caused by the interaction, users need to integrate multiple sensory information related to the change, which is obtained in visual, auditory, tactile, and kinesthetic modalities. To determine whether or not the information

should be integrated, sensory mechanisms in users employ various stimulus factors. For instance, users integrate the inputs from multisensory modalities based on their temporal proximity [27], [28]. Temporal discrepancies between the sensory inputs can result in the biased estimation of the object's characteristics [29], [30], which is often formulated via maximum likelihood estimation [31].

Previous studies in the field of human-computer interaction have proposed some methods to provide rich sensory experiences and/or improve manipulation performances by applying the characteristics of the human multisensory integration (e.g., [32]–[36]). It is known that the presentation of haptic impressions is modulated by the temporal integration between auditory and tactile information [29], which indicates that auditory stimuli are one of the valid sources for the modulation of haptic impressions in users [37], [38]. In this respect, it is worth checking how auditory stimuli can modulate the heaviness sensation.

2) VISUALLY INDUCED HEAVINESS SENSATION

While the present study proposes a method for modulating a sense of heaviness through the manipulation of auditory feedback in response to a user's action in a multimodal manner, most previous studies have proposed methods that rely on the manipulation of visual feedback after an action has been performed. Conventional methods can be classified into two categories: those that manipulate the ratio of visualized displacement to input displacement, and those that manipulate delay.

The former is the ratio of the amount of an avatar's movement, such as that of a cursor on a monitor, to the amount of the user's input movement, such as that made using a mouse. The larger the ratio, the greater the sense of heaviness the user experiences. In earlier studies, users who grasped and lifted a real object have reported a greater sense of heaviness when the visual speed of the object is reduced (i.e., when the ratio is larger) [39]–[41].

Independently of pseudo-haptic techniques in human-computer interaction, delayed visual feedback has been a focus of attention in the field of motor control, as it is a parameter that can induce an illusory sense of force such as heaviness or resistance. For example, Honda *et al.* reported that when a participant's arm movement was linked to the motion of a virtual object as visual feedback, a 200-400 ms delay in the visual feedback resulted in a greater perception of mass than without that delay [10]. In another study, when participants periodically flexed and extended their wrist while seeing an image of their hand with a delay of 150-600 ms, they reported that the wrist seemed to become heavier [11]. It has been proposed that the illusory force sensation produced by the delayed visual feedback results from the brain's reselection of an internal model of the mechanical load applied to the hand, rather than from an error itself between the predicted and the visually observed hand position [13].

3) PERCEPTION OF DELAYED SOUND FEEDBACK IN RESPONSE TO USER ACTION

In this study, we propose a method using auditory feedback, such as a beep with a certain delay applied to it, that is provided in response to a mouse button click. The perception of delayed auditory feedback in response to a user's action has been thoroughly investigated in studies into the human delay detection threshold and causal perception.

To investigate the delay detection threshold, researchers conducted experiments in which they asked participants to judge which of the events came first, that is, their action (e.g., a mouse click) or the sound feedback, while changing the temporal interval between the events [42], [43]. It is known that detection thresholds for a delay in action feedback vary within the range of delays employed in stimulus sets. For example, when the delay ranged between 0 and 200 ms, the delay detection threshold was 40-50 ms [42], [43]. On the other hand, the wider the range of the delay used in the experiment, the larger the delay detection range tended to be. For example, when the delay ranged between 0 and 419 ms, the delay detection threshold was about 300 ms [44]. The delay detection threshold is also altered by adaption to a specific delay [42], [43].

The perceived causality between a user's keypress (or physical button press) and corresponding auditory feedback is also weakened when a delay is added to the auditory feedback. It has been shown that participants tended to perceive the timing of the keypress and the timing of the beep presentation as being closer to each other than they actually were [15]. This effect becomes weaker as the time interval between the keypress and the beep increases (specifically, when the delay range was in the 250-650 ms range) [15]. These results suggest that subjective proximity governs the strength of subjective causality between the keypress and the auditory feedback. Consistent with this suggestion, when participants are asked to evaluate the sense of agency they feel (i.e., the sensation that they caused the beep), it has been shown that a 100-600 ms delay between the keypress (or physical button press) and the beep weakened the sense of agency [16], [17].

Although previous studies have investigated the perception of delay in auditory feedback for a user's action and its effect on the sense of agency, no study has examined the effect of the delay in auditory feedback on the sense of heaviness.

4) HAPTIC ILLUSION WITH SOUND FEEDBACK

The method proposed in this study modulates the auditory feedback provided in response to a user's action. It has been reported that the perception of tactile texture can be changed by modulating the auditory feedback provided in response to the rubbing of textured surfaces; this is called the parchment skin illusion. In this illusion, enhanced high-frequency feedback of the sound of two palms rubbing together made the palms feel drier and more like parchment paper [45]. This kind of audio-haptic interaction arises possibly because

temporal frequency channels are linked between the two sensory modalities [46]. The illusion is weakened when the auditory feedback is delayed [47], which suggests that temporal consistency between an individual's action and sound feedback is important for the illusion to occur. These findings show the possibility that humans have the ability to judge haptic sensations based on auditory information related to self-motion, which supports the effectiveness of our method using action-related auditory feedback.

The illusion is not limited to the user rubbing his or her palms together but occurs when users judged the texture roughness of other objects. Specifically, it is known that white noise as auditory feedback alters the perception of tactile roughness [48]. Moreover, in a pilot study that examined pseudo-haptics using auditory feedback instead of conventional visual feedback, it was shown that the presentation of a white noise sound while a mouse cursor was located within a certain area could create a different subjective impression from that experienced when the mouse entered an area without a white noise sound being triggered [49]. Although this study did not report what impressions were produced or how strongly, their results suggest that pseudo-haptics can be induced by using auditory feedback.

5) SOUND-BASED RECOGNITION OF MATERIALS

In our method, we manipulate not only the delay in the auditory feedback in response to a user's button click but also the frequency and loudness of the sound. Even when listening to sounds passively without acting, humans can use the frequency and loudness of the sound to estimate the heaviness of an object. Specifically, lower-pitched and louder collision sounds were judged as being caused by larger and heavier objects [23]. The results suggest that humans associate lower-pitched or louder sounds with weight. Consistent with this finding is an experiment in which participants listened to sounds presented by a speaker and evaluated their impressions, in which participants tended to rate lower-pitched sounds as heavier [24].

This perceptual association between weight and sound exists not only when passively hearing a sound but also when actually touching an object. When grasping or lifting heavier objects, participants tended to associate a lower-pitched sound with a heavier object (note that no actual sound was presented) [50]. It was also reported that when a low-pitch sound was presented when a paper box was placed in the palm of a participant's hand, they rated the weight of the box as heavier than when a high-pitched sound was presented [26]. On the other hand, changing the loudness of the sound did not affect the weight evaluation [26]. A study examining the effect of auditory feedback on performance when picking a virtual object noted an introspective report that the user interpreted the object to be heavier when the sound pitch played during the object picking was lower [51]. Another study reported that the weight of one's avatar, although not an external object, was rated heavier as the central frequency of the avatar's footsteps was lower [52].

These findings suggest that humans have, in their daily lives, empirically and statistically associated the properties of the sounds produced by an object with the heaviness of that object. Although these earlier studies did not examine how the frequency and loudness of auditory feedback in response to a user's action modulated the user's estimation of heaviness, we expect that the association between object heaviness and sound frequency and loudness may allow us to modulate the user's estimation of heaviness by presenting a lower-pitched or louder action-related sound.

C. OVERVIEW

The purpose of this study was to investigate how the modulation of auditory feedback in response to a user's button click altered the sense of heaviness perceived by the user. As shown in Fig. 1, we used a situation in which a button was displayed on the screen and the user clicked on the button using a mouse. After the button was clicked, a pure tone was presented as auditory feedback through headphones or speakers. The button click did not alter the appearance of the button and hence no visual feedback was given to the experiment participants. We aimed at modulating the subjective magnitude of the heaviness by controlling the delay between the button click and the presentation of the pure tone, the frequency of the pure tone, and the loudness of the pure tone.
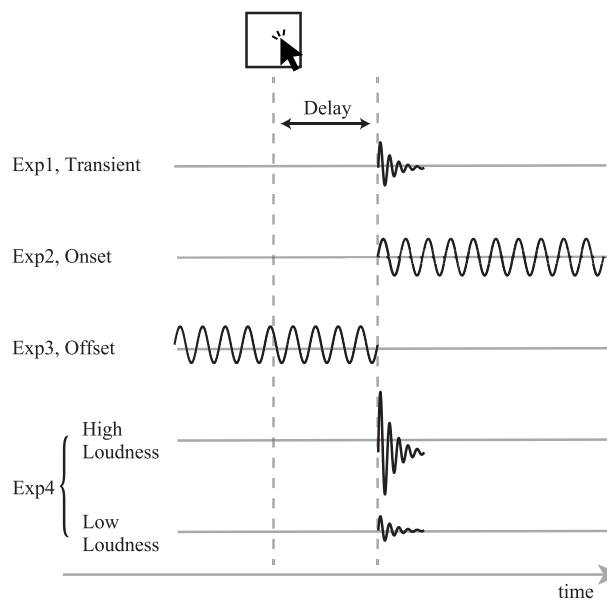
FIGURE 1. Temporal design of feedback sound stimuli in experiments 1-4. Note that, in addition to the control of delay and loudness of the sound as shown in this figure, we also manipulated the frequency of the sound in Experiments 1-3.

We hypothesized that the sense of heaviness would increase with the delay between the button click and the auditory feedback, with a lowering of the tone frequency, or with an increase in the tone loudness. To test these hypotheses, we conducted four experiments.

In Experiment 1 (section II), we examined whether a delay between a button click and the presentation of a transient pure tone would affect the subjective heaviness evaluation (Fig. 1a). The effect was compared with that of the frequency of a pure tone on the heaviness modulation. In Experiment 2 (section III), we tested whether a delay in the onset of the feedback tone affected the illusory sense of heaviness: when the user clicked a button, the tone was presented for a longer duration than that used in Experiment 1 (Fig. 1b). In Experiment 3 (section IV), we tested whether a delay in the offset of the feedback tone affected the illusory sense of heaviness: the pure tone was presented from the beginning of each trial, and when the button was clicked, the tone stopped after a certain delay (Fig. 1c). We again compared the effect of delay in tone onset and offset to the effect of the tone frequency. Note that, in Experiments 1, 2, and 3, we asked the participants to adjust the perceived loudness to be the same for each sound frequency condition to separate the effect of sound frequency from the difference in physical intensity of the sound. Finally, in Experiment 4 (section V), we examined how the loudness and delay of the feedback tone affected the heaviness evaluation (Fig. 1a).

II. EXPERIMENT 1: EFFECTS OF DELAY AND FREQUENCY OF FEEDBACK SOUND

A. PURPOSE

The purpose of the experiment was to examine whether the participants could feel heaviness in relation to the temporal delay and frequency of a pure tone triggered by their button clicks. In the experiment, the participants were asked to click a button on a display to play a pure tone. The delay in the onset of the pure tone from the time point of the participant's button click was varied between five different intervals: 0, 100, 200, 300, and 500 msec. Also, the frequency of the pure tone was varied between two different frequency levels (low-pitch: 200 Hz and high-pitch: 400 Hz) and it was compared to a pure tone of medium frequency (300 Hz) as a reference stimulus. The effects of delay and frequency on the degree of perceived heaviness were tested.

B. MATERIALS AND METHODS

1) PARTICIPANTS

One hundred and thirty-one people (65 females and 66 males) participated in the experiment and the mean ± standard deviation (SD) of their age was 40.27 ± 11.18 years. By using a statistical calculator, Morepower 6.0 [53], we calculated the sample size for a within-subjects design: a medium-effect size (Cohen's F = 0.25), a power of 80%, and an alpha of 5%. Minimum sample size that satisfies these conditions for all factors was 126. Therefore, we recruited participants so that the sample size would be more than 126. A Japanese crowdsourcing research company recruited the participants online and they were paid for their participation. They were unaware of the specific purpose of the experiment. Ethical approval for this study was obtained from the ethics committee at Nippon Telegraph and Telephone Corporation

(Approval number: R02-009 by NTT Communication Science Laboratories Ethics Committee). The experiments were conducted according to principles that have their origin in the Helsinki Declaration. Written informed consent was digitally obtained from all observers in this study.

2) APPARATUS

The experiment conducted in this study was carried out using the participants' own personal computers (PC) because our experimental script could only be run on a PC. Hence, smartphones or tablet PCs, which do not have keyboards, could not be used in this experiment. Viewing distance and screen size were not controlled because their effect was not evident in the preliminary observation provided the user used the PC normally.

3) STIMULI

The auditory feedback presented after a button click was a sinusoidal pure tone that decayed according to the following equation:

$$Q(t) = Ae^{-Bt}\sin(2\pi\omega t) \tag{1}$$

In the above equation, the amplitude, A, was defined as the maximum amplitude that is allowed in a wav file, and the decay rate, B, was defined as 20. The sampling frequency was defined as 44 kHz. The duration of the feedback sound was 500 ms. We adopted the pure sinusoidal tone with temporal decay because it is known to be a simple and natural model for haptic and auditory feedback when tapping a virtual object [54].

As shown in Figure 2a, the visual stimuli consisted of two square buttons each having sides of 100 × 100 pixels and a grayscale value of 255. The buttons were presented side by side in the center of the display against a uniform background (with a grayscale value of 128). The square buttons were sequentially presented: the first one was presented on the left side of the display and the second one on the right side of the display. When a participant clicked each button, a feedback tone was played with a delay. No visual change occurred in the button. Clicking one of the buttons triggered either a reference or comparison stimulus. Button assignment to the reference or comparison stimulus was randomly determined for each trial. The reference stimulus had a 300 Hz frequency and was always presented with a 200 ms delay after the participant's click. The comparison stimulus had one of two frequencies (200 Hz or 400 Hz) and was accompanied by one of five delays (0, 100, 200, 300, and 500 ms). The delay of the tone was determined on the basis of the time elapsed from the participant's click, not on the elapsed number of frames. Therefore, the delay accuracy was not substantially affected by the frame rate of a participant's PC. In addition to normal trials, we conducted catch trials to ascertain whether the participants gave appropriate responses. In the catch trials, no sound was played when participants clicked the first and second buttons.

In order to check the latency between the participants' mouse clicks and the feedback sound being emitted in

an online experimental environment, the temporal interval between the onset of a mouse click and the sound emitted from the speaker of the author's PC was preliminarily measured as "actual delay". The actual delay was measured twenty times for each delay condition (i.e., five delay conditions × 20 repetitions) in random order. The results are shown in Fig 2b. The result of regression analysis for the actual delay as a function of expected delay, was a slope of 1.00, with an intercept of 102.12. The slope of 1.00 meant that the system delay was constant across the delay conditions we tested. Therefore, for all delay conditions, the relative differences between the expected and actual delays were kept constant at approximately 102 ms.

4) PROCEDURE

At the beginning of the experiment, the participants were presented with written instructions that described the situation and their tasks in the experiment. After reading this, the participants adjusted the volume of the pure tone of 300 Hz (i.e., reference stimulus) to a comfortable loudness. After that, participants were also asked to adjust the volumes of pure tones of 200 Hz and 400 Hz so that the intensity felt equal to that of the 300 Hz tone. The participants used the up and down keys to change the volume. After the volume adjustment, the experiment was started.

The participant's task was to click two square buttons sequentially presented on the display to trigger a sound. The first square button was presented on the left side of the display. After participants clicked the button, the first feedback tone was played with a delay. After 500 ms from the offset of the tone, the second square button was presented on the right side of the display. When participants clicked the button, the second feedback tone was played with a delay (note that the reference and comparison stimuli were randomly assigned to the first and second buttons). When 500 ms had elapsed from the offset of the second tone, the answer screen was presented. On the answer screen, the following instruction was shown to the participants, "Please rate which button was heavier on a scale of 1 to 5" Participants reported this on a 5-point scale by pressing the assigned keys (1: the left one was much heavier, 2: the left one was slightly heavier, 3: they were comparable, 4: the right one was slightly heavier and 5: the right one was much heavier). A key of 0 ("no sound was presented") was also presented as a choice option, on the assumption that participants would choose this option in catch trials in which no sound was presented. After reporting the evaluation, the next trial began.

We randomized the order of presentation every 11 trials (i.e., 2 frequency conditions × 5 delay conditions + 1 catch trial) and repeated them four times, thus, each participant performed 44 trials in total.

5) ANALYSIS

Before the analysis, we excluded participants who did not select the "no sound" option in the catch trial at least twice or who selected the "no sound" option in normal experimental trials at least once since it is highly likely that the experimental stimuli were not presented appropriately to them. Two participants were excluded by this procedure, and data from the other 129 participants were used for subsequent analyses.

Since it was randomly determined whether the comparison stimulus was assigned to the first (left) or second (right) button, the rating scores were sometimes reversed between these cases. Therefore, we inverted the rating scores in the former case so that the rating scores became higher when participants evaluated the comparison stimulus as heavier.

Rating scores using the Likert scale have upper and lower limits and do not exhibit normality under the conditions in which mean scores are located near the upper or lower limit. Therefore, we first carried out an aligned rank transform (ART) [55] for the rating scores, and then, conducted a two-way repeated measures analysis of variance (ANOVA) with the delay and the frequency as within-subject factors. Multiple tests with Bonferroni correction were then performed for the delay factor.

C. RESULTS AND DISCUSSION

Figure 3 shows the results of the heaviness ratings scores when the frequency and delay were changed. In this and the following plots, higher rating scores than 3 meant that the comparison stimulus was reported to be heavier than the reference stimulus (having frequency of 300 Hz and delay of 200ms). The ART-ANOVA test showed significant main effects of the delay condition ($F(4, 512) = 4.87, p < 0.001, \eta_p^2 = 0.04$). Multiple comparison tests for the delay conditions showed that the mean rating scores for the 0 ms and 100 ms delay conditions were significantly smaller than that for the 500 ms delay condition. The ART-ANOVA test also showed the significant main effect of the frequency condition ($F(1, 128) = 89.05, p < 0.001, \eta_p^2 = 0.41$), which means the significant difference between the 200 Hz and 400 Hz conditions (since there were only two conditions). The interaction between the delay and frequency was not significant ($F(4, 512) = 0.87, p = 0.48, \eta_p^2 = 0.01$).

The results showed that participants tended to report a stronger heaviness sensation when a lower-pitched feedback sound was presented after clicking the button. The results that the past finding that lower-pitched sounds induced a stronger heaviness sensation in passive situations [24] can be replicated even in an interactive situation.

We also found that participants tended to report a stronger heaviness sensation when the auditory feedback in response to their button click was presented with a longer delay. These results suggest that the illusory heaviness sensation which is induced by the conventional method of delayed visual feedback [10], [11] can also be induced by delayed auditory feedback. A similar mechanism which is independent of the sensory modalities might underlie the illusory heaviness sensation.

While we showed that presenting a transient sound with a delay could induce an illusory heaviness sensation, in our daily lives, the auditory feedback of clicking/pressing
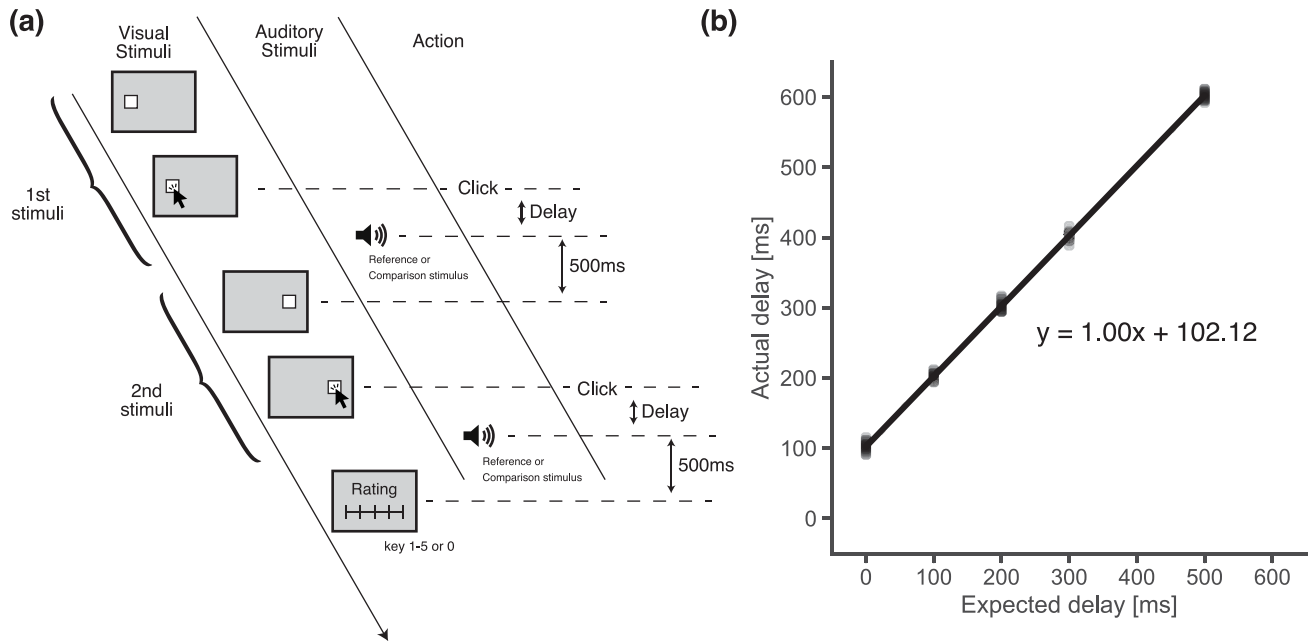
FIGURE 2. (a) Trial sequence of experiment 1. The presentation order of the reference and comparison stimuli was determined in a randomized manner for each trial. (b) The actual delay as a function of the expected delay, which was calculated to check the validity of our experimental program in an online experimental environment. Each circle represents the actual delay for each trial. The line represents a linear function regressed for the actual delay as a function of the expected delay.

a button is not always a transient sound. For example, when we turn on a radio by pressing a button, a sound will start to play, or a sound that had been playing will stop. It is important to investigate whether a delay in either the onset timing or the offset timing of the feedback sound can modulate the heaviness sensation in order to understand the applicability of our method and the mechanism underlying it. Since the transient sound used in this experiment had a fixed duration of 500ms, the onset and offset timings changed in the same way across different delay conditions. Therefore, in the subsequent experiments, we used a continuous sound instead of a transient sound to control the onset and offset timings independently and confirm the effect of the delay.

III. EXPERIMENT 2: EFFECT OF THE ONSET OF FEEDBACK SOUND

The purpose of the experiment was to examine whether the participants used the onset timing of delayed auditory feedback to estimate the heaviness sensation. In experiment 1, we observed that temporal delay of a transient feedback sound had an effect on the heaviness evaluation. In this experiment, we asked participants to click a button on a display to start playing a continuous pure tone. The delay in the onset of the pure tone feedback from the timing of the participant's button click was varied among five different intervals: 0, 100, 200, 300, and 500 msec. The other conditions and procedures were the same as in Experiment 1.

A. METHODS
1) PARTICIPANTS
One hundred and twenty-nine people (63 females and 66 males), who had not participated in Experiment 1,

participated in Experiment 2. The mean $\pm$ SD of their age was 39.79 ± 11.45 years. The protocol for the consent, recruitment, and ethics were identical those as used in Experiment 1.

2) STIMULI
The stimuli used in this experiment were identical to those in Experiment 1, except for the feedback sound presented after the button click. A continuous sinusoidal wave was used as the auditory feedback. When a participant clicked each button, a feedback sound was played at a constant volume without decay. As in Experiment 1, the reference stimulus had a 300 Hz frequency and was presented with a 200 ms delay after the button click. The comparison stimulus had one of two frequencies (200 Hz and 400 Hz) and was accompanied by one of five onset delays (0, 100, 200, 300, and 500 ms).

3) PROCEDURE
The procedure was identical to that of Experiment 1 except for the following points. In this experiment, the feedback sound kept playing after the button was clicked. Five hundred milliseconds after the onset of the feedback sound, the instruction to proceed to the next screen by pressing the Q key was displayed. When the Q key was pressed, the sound stopped and the next screen was displayed.

4) ANALYSIS
The analysis method was the same as in Experiment 1. We excluded six participants who did not select the "no sound" option in the catch trial at least twice or who selected the "no sound" option in normal experimental trials
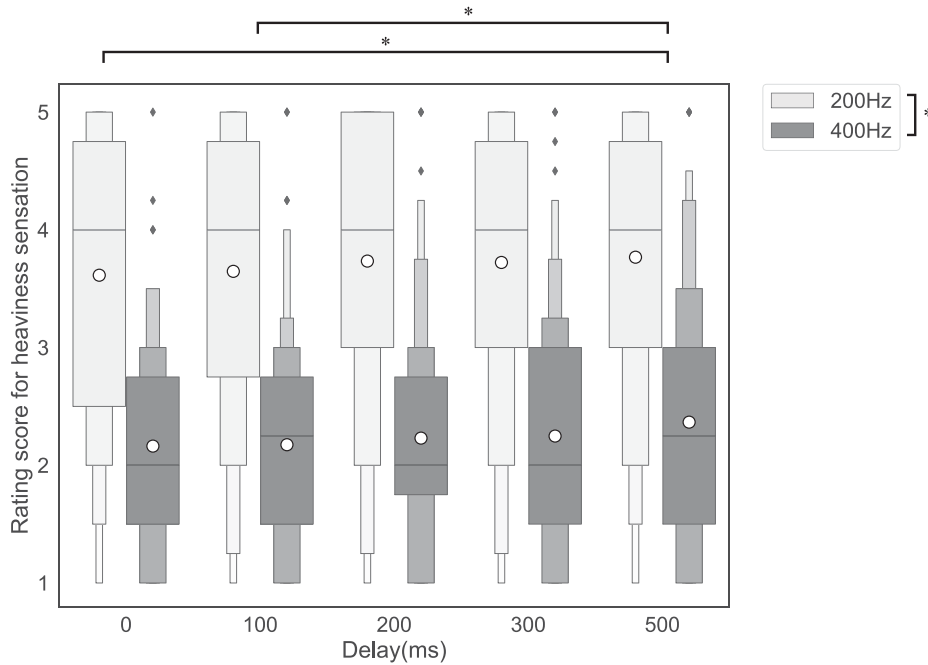
FIGURE 3. Letter-value plots of heaviness rating scores in experiment 1. The white dots and the horizontal lines indicate the mean and the median of the rating scores, respectively.

TABLE 1. Results of ART-ANOVA [55] in experiment 1.

Factor	df	df res	Sum of Squares	Mean Square	F	p	η_p^2
Frequency	1	128	52714352.28	52714352.28	89.05	0.00	0.41
Delay	4	512	655060.47	163765.12	4.87	0.00	0.04
Frequency x Delay	4	512	81043.44	20260.86	0.87	0.48	0.01

at least once. Thus, data from the other 123 participants were used for subsequent analyses.

B. RESULTS AND DISCUSSION

Figure 4 shows the results of the heaviness ratings scores when the frequency and delay onset were changed. The ART-ANOVA test showed the significant main effect of the frequency condition ($F(1, 122) = 126.58, p < 0.001, \eta_p^2 = 0.51$), which means the significant difference between the 200 Hz and 400 Hz conditions. On the other hand, the ART-ANOVA test did not show the significant main effect of the delay condition ($F(4, 488) = 1.24, p = 0.29, \eta_p^2 = 0.01$). The interaction between them was not significant ($F(4, 488) = 0.57, p = 0.68, \eta_p^2 < 0.01$).

In contrast to Experiment 1, the results showed that the heaviness perception was not affected by the onset of the delayed auditory feedback, even though the onset of the feedback sound should have been a powerful temporal marker to judge the presentation timing. A key difference in stimuli between Experiment 1 and the present experiment was the duration of the feedback sound. In Experiment 1, because the duration of the sound was short, the participants were able to use the offset as well as the onset of the feedback sound to

judge the heaviness. Since it turned out that the onset of the sound when a continuous sound was used did not play a major role in the determination of the heaviness sensation, there is a possibility that the illusory heaviness sensation arises from the delay of the offset rather than that of the onset of the feedback sound provided in response to a button click. To test this possibility, we conducted the next experiment wherein the offset timing of the sound was manipulated.

IV. EXPERIMENT 3: EFFECT OF THE OFFSET OF FEEDBACK SOUND

The purpose of the experiment was to examine whether the offset timing of the feedback sound contributed to the heaviness sensation. The heaviness sensation evoked with a transient pure tone in Experiment 1 was not evoked with a continuous tone in Experiment 2. A possible hypothesis derived from these results is that the offset, rather than the onset, plays a major role in the heaviness sensation. To address this issue, we conducted Experiment 3, wherein the pure tone was presented from the beginning of each trial, and when the button was clicked, the tone stopped after a delay. The delay of the offset of the pure tone from the timing of the participant's button click was varied among
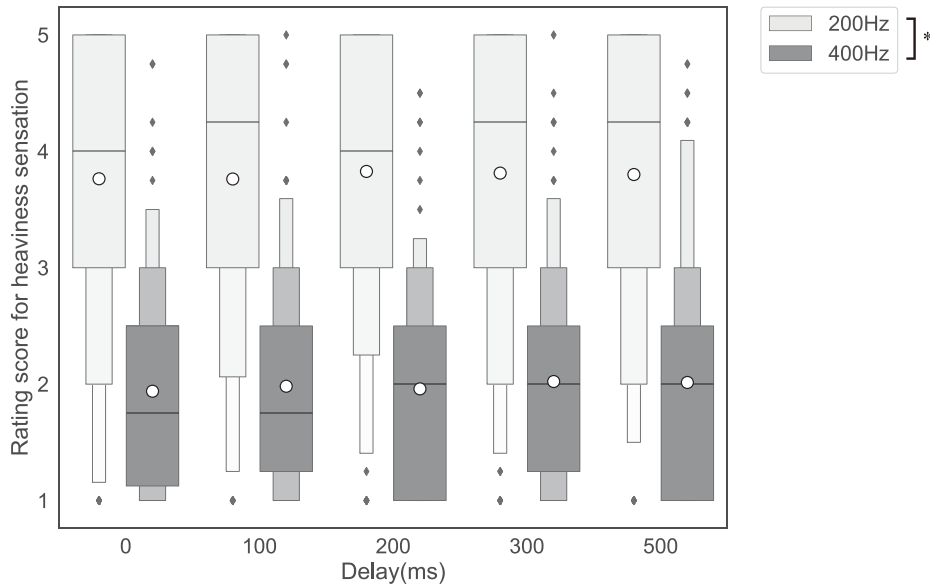
FIGURE 4. Letter-value plots of heaviness rating scores in experiment 2. The white dots and the horizontal lines indicate the mean and the median of the rating scores, respectively.

TABLE 2. Results of ART-ANOVA [55] in experiment 2.

Factor	df	df res	Sum of Squares	Mean Square	F	p	η_p^2
Frequency	1	122	59328454.28	59328454.28	126.58	0.00	0.51
Delay	4	488	92506.55	23126.64	1.24	0.29	0.01
Frequency x Delay	4	488	35817.35	8954.34	0.57	0.68	0.00

five different levels: 0, 100, 200, 300, and 500 ms. The other conditions and procedures were identical to those used in Experiment 1.

A. METHODS

1) PARTICIPANTS

One hundred and thirty people (64 females and 66 males), who had not participated in Experiments 1 and 2, participated in Experiment 3. The mean ± SD of their age was 39.91 ± 11.48 years. The protocol for the consent, recruitment, and ethics were identical those as used in Experiment 1.

2) STIMULI

The stimuli presented were identical to those in Experiment 1, except for the way of providing feedback after the button click of the participants. Similarly to Experiment 2, a pure tone with a sinusoidal wave was used as auditory feedback. In contrast to Experiments 1 and 2, the sound was played from the beginning of each trial and stopped with some delay when the participant clicked the button. As in Experiment 1, the reference stimulus had a 300 Hz frequency and was presented with a 200 ms delay after the button click. The comparison stimulus had one of two frequencies (200 Hz and 400 Hz) and was accompanied by one of five delayed offsets (0, 100, 200, 300, and 500 ms).

3) PROCEDURE

The procedure was identical to that as used in Experiment 1 except for the following points. The task of the participant was to listen to the sound that was played from the beginning of each trial and stop it by clicking the button on the monitor. Five hundred milliseconds after the end of the sound, the participants were allowed to press the Q key to display the second button (in the case of the first button click) or the answer screen (in the case of the second button click).

4) ANALYSIS

The analysis method was the same as in Experiment 1. We excluded seven participants who did not select the "no sound" option in the catch trial at least twice or who selected the "no sound" option in normal experimental trials at least once. Thus, data from the other 122 participants were used for subsequent analyses.

B. RESULTS AND DISCUSSION

Figure 5 shows the results of the heaviness ratings scores when the frequency and delayed offset were changed. The ART-ANOVA test showed significant main effects of the delay condition ($F(4, 484) = 7.09, p < 0.001, \eta_p^2 = 0.06$). Multiple comparison tests showed that the mean rating score for the 0 ms delay condition was significantly smaller than
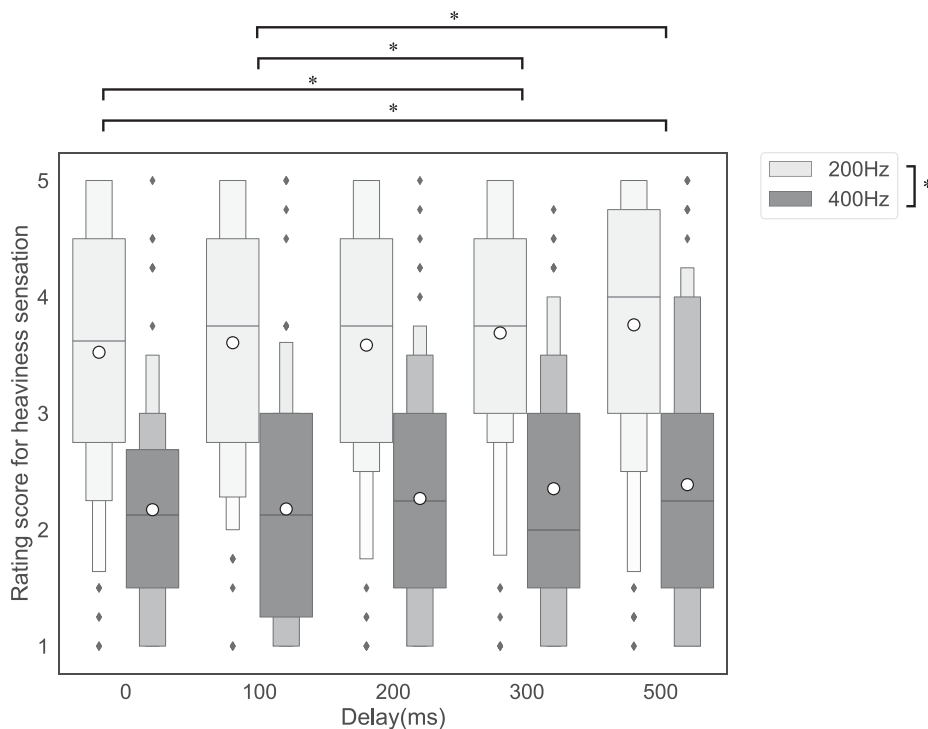
FIGURE 5. Letter-value plots of heaviness rating scores in experiment 3. The white dots and the horizontal line indicate the mean and the median of the rating scores, respectively.

that for the 300 ms and 500 ms delay conditions and that the mean rating score for the 100 ms delay condition was significantly smaller than that for the 300 ms and 500 ms delay conditions. The ART-ANOVA test also showed the significant main effect of the frequency condition ($F(1, 121) = 87.43, p < 0.001, \eta_p^2 = 0.42$), which means the significant difference between the 200 Hz and 400 Hz conditions. The interaction between them was not significant ($F(4, 484) = 0.70, p = 0.59, \eta_p^2 < 0.01$).

While the manipulation of onset timing in Experiment 2 did not affect the heaviness evaluation, the manipulation of offset timing in this experiment did affect it. The results suggest that the offset timing of the feedback sound plays a critical role in evoking a sense of heaviness accompanying the button click.

V. EXPERIMENT 4: EFFECTS OF LOUDNESS OF FEEDBACK SOUND

The purpose of the experiment was to examine the effect of the loudness of a transient pure tone accompanying a button click on the illusory heaviness sensation as well as the effect of delay. The experimental procedures were identical to those used in Experiment 1, except that we manipulated the sound loudness condition instead of the sound frequency condition. The loudness of the pure tone was varied between two different levels (low and high) and it was compared to reference stimuli of a pure tone with medium volume. The effects of loudness and delay on the illusory heaviness sensation were tested.

A. METHODS

1) PARTICIPANTS

One hundred and twenty-seven people (63 females and 64 males), who had not participated in Experiment 1, 2, or 3, participated in Experiment 4. The mean ± SD of their age was 39.75 ± 11.52 years. The protocol for the consent, recruitment, and ethics were identical those as used in Experiment 1.

2) STIMULI

The stimuli presented were identical to those in Experiment 1, except that we manipulated the loudness of the feedback sound instead of the frequency. We used the same decaying sine wave as in Experiment 1. The comparison stimulus has two levels of loudness (low and high) and five delay conditions (0, 100, 200, 300, and 500 ms). The volumes of the low and high conditions were set by the participants before the start of the experiment (see next section on procedure) as the minimum volume that could be heard and the maximum volume that was not considered too loud, respectively. The reference stimulus was the mean volume of the low and high loudness conditions, and was emitted with a 200 ms delay after the button click. The frequency of both the reference and comparison stimuli was 300 Hz.

3) PROCEDURE

The procedure was the same as that in Experiment 1 except for the following points. After reading the instructions, the participants were asked to adjust the volume of a pure tone

TABLE 3. Results of ART-ANOVA [55] in experiment 3.

Factor	df	df res	Sum of Squares	Mean Square	F	p	η_p^2
Frequency	1	121	45954728.20	45954728.20	87.43	0.00	0.42
Delay	4	484	912013.21	228003.30	7.09	0.00	0.06
Frequency x Delay	4	484	63592.97	15898.24	0.70	0.59	0.01

of 300 Hz. To determine the volume of the comparison stimulus in our "Low" loudness condition, they were asked to turn the volume down to the point that it was barely audible. After that, to determine the volume of the comparison stimulus in our "high" loudness condition, they were asked to turn the volume up to the point where it was loud but not too loud. To change the volume, the participants used the up and down keys. After the volume adjustment, the experiment was started.

4) ANALYSIS

In this experiment, we adopted an additional criterion for excluding participants who erroneously adjusted the volumes of stimuli in the small and large loudness conditions to be equal or inverted since this implied that the loudness of the feedback stimuli presented to the participants was not appropriate for our experimental purpose. First, we excluded 67 participants based on this criterion. Next, in the same manner as in Experiment 1, we also excluded three other participants who did not select the "no sound" option in the catch trial at least twice or who selected the "no sound" option in normal experimental trials at least once. Thus, data from the other 57 participants (29 females and 28 males; mean ± SD of their age was 39.26 ± 11.84 years) were used for subsequent analyses. Note that even after excluding these participants, the gender ratio and the mean and SD of their ages were almost the same as before the exclusion.

B. RESULTS AND DISCUSSION

Figure 6 shows the results of the heaviness ratings scores as a function of loudness and delay. The ART-ANOVA test showed a significant main effect of the loudness condition ($F(1, 56) = 17.82, p < 0.001, \eta_p^2 = 0.24$), which means the significant difference between low and high loudness conditions. We also found a significant main effect of the delay condition ($F(4, 224) = 15.85, p < 0.001, \eta_p^2 = 0.22$). Multiple comparison tests showed that the mean rating score for the 0 ms delay condition was significantly smaller than that for the 200, 300, and 500 ms delay conditions and that the mean rating score for the 500 ms delay condition was significantly larger than that for the 100, 200, and 300 ms delay conditions. The interaction between them was not significant ($F(4, 224) = 0.22, p < 0.93, \eta_p^2 < 0.01$).

The results showed that the feedback sounds accompanying the participants' button clicks that produced the stronger sensation of heaviness were the louder sounds. The participants might have estimated the heaviness influenced by their prior knowledge that heavier objects tend to produce a louder

sound upon impact [23]. Contrary to our results, a previous study [26] that manipulated the loudness of the sound presented when a box was placed in the palm of the hand reported no significant effect of the loudness on the heaviness judgement. The effect of loudness in our study might be unique to the situation in which users produced the feedback sound by performing an action, in our experiment by clicking a button. This hypothesis that the correspondence between an action and the feedback is important for the loudness effect could be tested by directly comparing the active condition (i.e., the results in this study) with the passive condition (i.e., the results of [26]) in the similar experimental setting.

We also replicated the effect of delay in auditory feedback on the illusory heaviness sensation which occurred in Experiment 1. However, the effect size of the delay in this experiment was much greater than that in Experiment 1. The difference in the effect sizes might be explained in terms of a relative contribution of delay and other cues to the heaviness sensation. The effect size for the main effect of frequency in Experiment 1 was greater than the effect size for the main effect of loudness in Experiment 4. The results indicate that the contribution that frequency makes to the sensation of heaviness is greater than that of loudness. When delay is experimentally paired with these cues, the relative contribution of delay is possibly greater when it is paired with loudness than when it is paired with frequency. Thus, though delay is an effective cue to the heaviness sensation as shown in this experiment, the effect of delay is possibly reduced when other strong cues such as frequency are presented.

VI. DISCUSSION
A. PERCEPTUAL FACTORS
Our results showed that participants tended to report a stronger sensation of heaviness when auditory feedback was presented with a longer delay. On the other hand, the mechanism of the illusory heaviness sensation remains unclear. As described above, the brain possibly internalizes the relationship between object heaviness and the frequency/loudness of sound [23], [24]. Similarly, the brain may internalize a statistical relationship between the heaviness of an object and the movement delay that occurs when a force is applied to that object. Our ancestors may have repeatedly experienced that heavy objects take longer to move after force is applied to them, and in the process of evolution, the brain has possibly internalized the statistical relationship between object heaviness and the delay. The results of the present experiment may indicate that due to those internalized statistics, the brain judges an object to be heavier when it
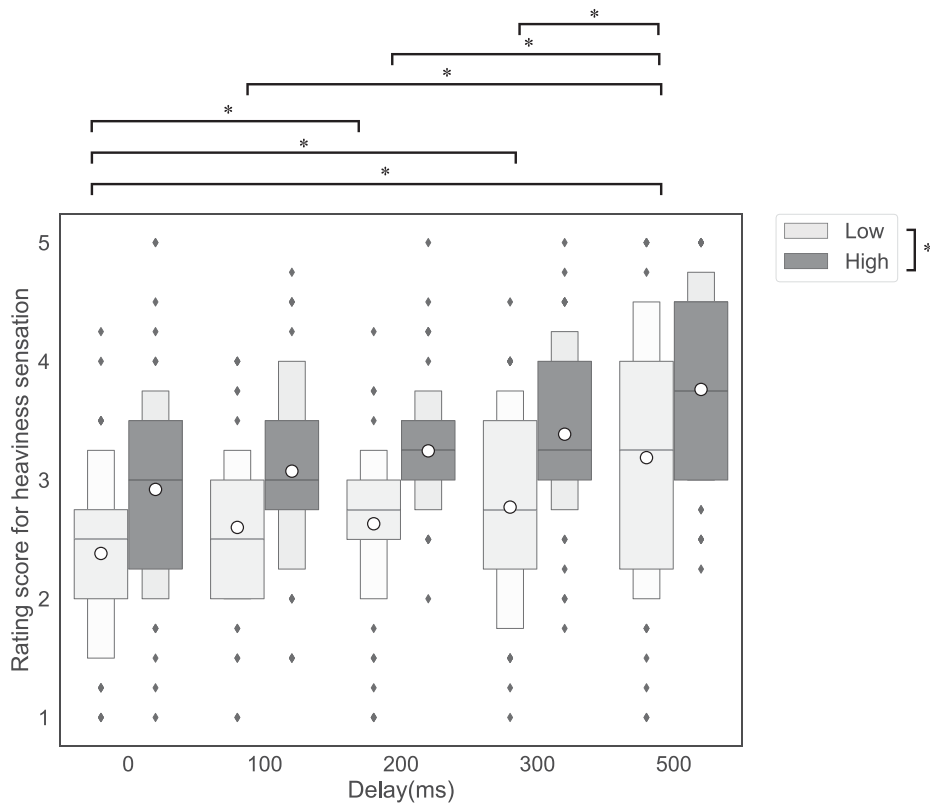
FIGURE 6. Letter-value plots of heaviness rating scores in experiment 4. The white dots and the horizontal line indicate the mean and the median of the rating scores, respectively.

TABLE 4. Results of ART-ANOVA [55] in experiment 4.

Factor	df	df res	Sum of Squares	Mean Square	F	p	η_p^2
Loudness	1	56	1697623.59	1697623.59	17.82	0.00	0.24
Delay	4	224	1462121.95	365530.49	15.85	0.00	0.22
Loudness x Delay	4	224	8244.12	2061.03	0.22	0.93	0.00

takes longer to start moving or otherwise changing when the human interacts with it.

The offset timing of auditory feedback affected the judgments of the heaviness sensation while the onset timing did not, which might suggest that participants were not able to correctly detect the delay of feedback sounds when the onset timing was varied. The results are interpreted in terms of 1) perceptual properties for the timing of the sound onset and 2) perceptual properties for the relative timing between the button click and the sound onset. For 1), it has been shown that the sound onset was perceptually more biased toward the offset as the sound duration increased [56]. In this respect, the onset timing of our stimuli as used in Experiment 2 might be perceptually biased toward the offset timing which was determined by the participants, and this might eliminate the effect of the delay on the heaviness sensation. For 2), a previous study [57] has demonstrated that cross-modal temporal order judgments were influenced by stimulus duration. Specifically, the sound onset timing in relation to visual onset

was more perceptually delayed when the sound duration was longer than when it was shorter. There is a possibility that in our experiment, the sound onset in relation to the timing of the button click was perceptually delayed under all delay conditions, and due to this, the effect of the delay on the heaviness sensation was attenuated.

We found that lower-pitched and louder feedback sounds induced a stronger heaviness sensation. The results are consistent with the physical properties of sounds in a natural scenario, where heavier objects tend to emit lower-pitched and louder sounds [23]. Based on comparing the effect sizes between Experiment 1 and Experiment 4, the effect of sound frequency seemed to be greater than that of sound loudness. The reason why the effect of loudness was relatively slight might be that the cause of the differences in sound loudness cannot be uniquely attributed to the weight of the object. In fact, loudness tends to be affected not only by the weight of the objects but also by other multiple factors such as the force applied to the object, the speed of its movement, and the user's

distance from it. On the other hand, the sound frequency is closely related to the material of the object [25], and it may be relatively easy to attribute the cause to the weight (i.e., the density of the material).

B. TECHNICAL FACTORS

Our method is a powerful means of inducing a sensation of heaviness using auditory stimuli. This idea is supported by the robust effects of delay and frequency which were observed across multiple experiments. It is, in a sense, surprising to obtain such robust effects in online experiments wherein different participants would be expected to have diverse listening environments (e.g., we did not know whether they used speakers or headphones, and it was also unclear whether they were participating in a quiet or noisy environment). Because of the simplicity of our experimental method, there is still some room to improve certain technical aspects.

While we used pure tones with low (200 Hz) and high (400 Hz) frequencies in this experiment, in practical use, pure tones having different frequencies from those we chose could be used as feedback sounds. Although we found that a low-frequency (200 Hz) pure tone induced a heavier sensation than the high-frequency (400 Hz) pure tone, it is an open question as to whether the relationship between the heaviness sensation and the tone frequency can be maintained for other frequency pairs (for example, 400 and 800 Hz). Another open question is whether it is necessary to add temporal decays to sound feedback. Although we used decaying sinusoidal waveform as the transient sounds in our study as observed in the real world, it is still unclear whether the real-world-like decay plays a significant role in inducing the illusory sensation of heaviness using temporal delay. In addition, when a user controls common information devices, the auditory feedback of the user's button click is not always presented in isolation; it will occasionally be presented with other sound information such as natural sounds, music, and voice. It is unclear whether the effect of delay on the heaviness sensation would still be observed even in noisy environments with other sound sources.

Based on the results of Experiment 1 showing that the delay of a transient sound after a user's button click caused the heaviness sensation, we believe that the application of our method is feasible in some types of devices where a transient sound is used as the auditory feedback of a user's action like a button click. On the other hand, in other types of devices such as radios, the sound being fed back after a user's action is more likely to be speech or music, the duration of which is usually longer and which has a considerably more complex temporal structure than that of pure sound. Because it was shown in Experiment 2 that the onset delay of long pure tones did not cause the heaviness sensation, the delay is unlikely to be effective in inducing a heaviness sensation when devices that use long duration sounds as feedback are employed. On the other hand, if the presence of the offset and

its timing is an important factor in inducing the heaviness sensation, inserting an explicit sound offset (temporal gap) into continuous sounds may allow us to modulate the heaviness sensation.

Our proposed method might be compatible with visually induced pseudo-haptics (i.e., extension of our method to multimodalities). For example, when presenting a visual delay and an auditory delay at the same time, users would be able to experience the pseudo-haptics effect even when they are not looking at the screen. Two issues need to be investigated in this direction. The first is whether the perceived heaviness increases when visual and auditory methods are used simultaneously, compared to when either of the methods is used. The second is whether the heaviness disappears when visually- and auditory-induced heaviness sensations are inconsistent. These issues will need to be addressed in conjunction with understanding the brain's computational processes of how visual and auditory information about heaviness is integrated.

While our purpose in this study was to investigate the averaged tendency across all participants, there might be several individual differences. For example, the strength of learning of the co-occurrence relationships of sound and heaviness, which would be important for estimating heaviness from sound, might vary from person to person and affect the intensity of the illusory heaviness sensation induced by sound. Thus, investigating whether the effectiveness of the proposed method changes based on the user's background and past experience is an important direction for future studies to assess the applicability of the method. Moreover, differences in the participants' listening environments (e.g., whether they used speakers or headphones and whether they were participating in a quiet or noisy environment) might result in the individual difference in our experiments. Since users are expected to operate their devices in a variety of listening environments, it is also important to investigate how different listening environments affect heaviness sensation.

Here, we summarize the pros and cons of using our method for presenting illusory heaviness sensation. The advantage of our method is that it requires only a very simple speaker and can be implemented on many devices. This advantage contributes to the miniaturization of the device, as it can generate illusory heaviness sensation without a device that requires a lot of space, such as a monitor. On the other hand, the difficulty in presenting spatial information is a disadvantage of our method. It is not easy to present auditory feedback according to each of the user's simultaneous inputs for multiple locations. Also, auditory feedback is inaudible in noisy environments, limiting the method's effectiveness.

The proposed method can be used under application scenarios wherein using visual feedback is inappropriate to cause the heaviness impression. One example for potential applications is a camera app of smartphones. If visual feedback is applied to a shutter button in the interface of the camera app, the visual perception of the camera image as well as the button of the camera app will be undesirably deterio-

rated. For the app of this sort, using auditory feedback is more suitable because no visual distraction occurs with the auditory feedback. In another direction, our method is also useful for presenting illusory heaviness on devices with a small screen on which sufficient visual information cannot be presented. In the applications of this direction, for example, the heaviness sensation can be induced by the auditory feedback for the operation of devices such as wearable activity trackers or wireless earphones. Implementing our proposed method for these possible applications and comparing the effects with the results in this study will help us understand how robustly the proposed method works in the various contexts of possible applications.

VII. CONCLUSION

We proposed a new method to induce a heaviness sensation by manipulating the delay, frequency, and loudness of auditory feedback associated with a user's input action. In a series of psychophysical experiments, we confirmed that these parameters effectively modulated the heaviness sensation. The results of our method suggest that the heaviness sensation, which has conventionally been induced by modulating visual feedback, can also be caused by modulating auditory feedback. This method can be further extended to present haptic properties to the user even in simple information devices without visual and/or tactile displays.

REFERENCES

[1] T. H. Massie and J. K. Salisbury, "The phantom haptic interface: A device for probing virtual objects," in *Proc. ASME Winter Annu. Meeting, Symp. Haptic Interfaces Virtual Environ. Teleoperataor Syst.*, 1994, vol. 55, no. 1, pp. 295–300.

[2] T. Endo, H. Kawasaki, T. Mouri, Y. Ishigure, H. Shimomura, M. Matsumura, and K. Koketsu, "Five-fingered haptic interface robot: HIRO III," *IEEE Trans. Haptics*, vol. 4, no. 1, pp. 14–27, Jan. 2011.

[3] F. Kimura, A. Yamamoto, and T. Higuchi, "Development of a 2-DOF softness feeling display for tactile tele-presentation of deformable surfaces," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 1822–1827.

[4] Y. Vardar, A. Isleyen, M. K. Saleem, and C. Basdogan, "Roughness perception of virtual textures displayed by electrovibration on touch screens," in *Proc. IEEE World Haptics Conf. (WHC)*, Jun. 2017, pp. 263–268.

[5] A. Lecuyer, S. Coquillart, A. Kheddar, P. Richard, and P. Coiffet, "Pseudo-haptic feedback: Can isometric input devices simulate force feedback?" in *Proc. IEEE Virtual Reality*, Mar. 2000, pp. 83–90.

[6] Y. Ujitoko and Y. Ban, "Survey of pseudo-haptics: Haptic feedback design and application proposals," *IEEE Trans. Haptics*, vol. 14, no. 4, pp. 699–711, Oct. 2021.

[7] L. Dominjon, A. Lecuyer, J. Burkhardt, P. Richard, and S. Richir, "Influence of control/display ratio on the perception of mass of manipulated objects in virtual environments," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2005, pp. 19–25.

[8] F. Argelaguet, D. A. G. Jáuregui, M. Marchal, and A. Lecuyer, "Elastic images: Perceiving local elasticity of images through a novel pseudo-haptic deformation effect," *ACM Trans. Appl. Perception*, vol. 10, no. 3, pp. 1–14, Aug. 2013.

[9] P. Punpongsanon, D. Iwai, and K. Sato, "*SoftAR*: Visually manipulating haptic softness perception in spatial augmented reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 11, pp. 1279–1288, Nov. 2015.

[10] T. Honda, N. Hagura, T. Yoshioka, and H. Imamizu, "Imposed visual feedback delay of an action changes mass perception based on the sensory prediction error," *Frontiers Psychol.*, vol. 4, p. 760, Oct. 2013.

[11] M. Osumi, S. Nobusako, T. Zama, M. Taniguchi, S. Shimada, and S. Morioka, "Sensorimotor incongruence alters limb perception and movement," *Hum. Movement Sci.*, vol. 57, pp. 251–257, Feb. 2018.

[12] M. Di Luca, B. Knörlein, M. O. Ernst, and M. Harders, "Effects of visual–haptic asynchronies and loading–unloading movements on compliance perception," *Brain Res. Bull.*, vol. 85, no. 5, pp. 245–259, Jun. 2011.

[13] S. Takamuku and H. Gomi, "What you feel is what you see: Inverse dynamics estimation underlies the resistive sensation of a delayed cursor," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 282, no. 1811, Jul. 2015, Art. no. 20150864.

[14] T. Yokosaka and T. Kawabe, "Delay and speed of visual feedback of a keystroke cause illusory heaviness and stiffness," *Frontiers Neurosci.*, vol. 16, Mar. 2022, Art. no. 761697.

[15] P. Haggard, S. Clark, and J. Kalogeras, "Voluntary action and conscious awareness," *Nature Neurosci.*, vol. 5, no. 4, pp. 382–385, Mar. 2002.

[16] S. Kühn, I. Nenchev, P. Haggard, M. Brass, J. Gallinat, and M. Voss, "Whodunnit? Electrophysiological correlates of agency judgements," *PLoS ONE*, vol. 6, no. 12, Dec. 2011, Art. no. e28657.

[17] A. Sato and A. Yasuda, "Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership," *Cognition*, vol. 94, no. 3, pp. 241–255, Jan. 2005.

[18] T. Asai and Y. Tanno, "The relationship between the sense of self-agency and schizotypal personality traits," *J. Motor Behav.*, vol. 39, no. 3, pp. 162–168, May 2007.

[19] C. Farrer, G. Valentin, and J. M. Hupé, "The time windows of the sense of agency," *Consciousness Cognition*, vol. 22, no. 4, pp. 1431–1441, Dec. 2013.

[20] T. Kawabe, W. Roseboom, and S. Nishida, "The sense of agency is action–effect causality perception based on cross-modal grouping," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 280, no. 1763, Jul. 2013, Art. no. 20130991.

[21] W. Wen, A. Yamashita, and H. Asama, "The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect," *Consciousness Cognition*, vol. 36, pp. 87–95, Nov. 2015.

[22] W. Wen, A. Yamashita, and H. Asama, "The sense of agency during continuous action: Performance is more important than action-feedback association," *PLoS ONE*, vol. 10, no. 4, Apr. 2015, Art. no. e0125226.

[23] M. Grassi, "Do we hear size or sound? Balls dropped on plates," *Perception Psychophys.*, vol. 67, no. 2, pp. 274–284, Feb. 2005.

[24] L. Walker, P. Walker, and B. Francis, "A common scheme for cross-sensory correspondences across stimulus domains," *Perception*, vol. 41, no. 10, pp. 1186–1192, Jan. 2012.

[25] R. Klatzky, D. Pai, and E. Krotkov, "Perception of material from contact sounds," *Presence*, vol. 9, no. 4, pp. 399–410, Aug. 2000.

[26] M. Takashima, "Perceived weight is affected by auditory pitch not loudness," *Perception*, vol. 47, no. 12, pp. 1196–1199, Oct. 2018.

[27] Y.-C. Chen and C. Spence, "Assessing the role of the 'unity assumption' on multisensory integration: A review," *Frontiers Psychol.*, vol. 8, p. 445, Mar. 2017.

[28] C. V. Parise, C. Spence, and M. O. Ernst, "When correlation implies causation in multisensory integration," *Current Biol.*, vol. 22, no. 1, pp. 46–49, Jan. 2012.

[29] V. Occelli, C. Spence, and M. Zampini, "Audiotactile interactions in temporal perception," *Psychonomic Bull. Rev.*, vol. 18, no. 3, pp. 429–454, Mar. 2011.

[30] C. Spence, "Shitsukan—The multisensory perception of quality," *Multisensory Res.*, vol. 33, no. 7, pp. 737–775, Aug. 2020.

[31] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, Jan. 2002.

[32] Z. Yuan, S. Chen, G. Ghinea, and G.-M. Muntean, "User quality of experience of mulsemedia applications," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 11, no. 1, pp. 1–19, Oct. 2014.

[33] E. Kruijff, A. Marquardt, C. Trepkowski, R. W. Lineman, A. Hinkenjann, J. Maiero, and B. E. Riecke, "On your feet! Enhancing vection in leaning-based interfaces through multisensory stimuli," in *Proc. Symp. Spatial User Interact.*, Oct. 2016, pp. 149–158.

[34] J. Maiero, E. Kruijff, A. Hinkenjann, and G. Ghinea, "ForceTab: Visuo-haptic interaction with a force-sensitive actuated tablet," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 169–174.

[35] E. Pezent, A. Israr, M. Samad, S. Robinson, P. Agarwal, H. Benko, and N. Colonnese, "Tasbi: Multisensory squeeze and vibrotactile wrist haptics for augmented and virtual reality," in *Proc. IEEE World Haptics Conf. (WHC)*, Jul. 2019, pp. 1–6.

[36] Y. Zhao, C. L. Bennett, H. Benko, E. Cutrell, C. Holz, M. R. Morris, and M. Sinclair, "Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–14.

[37] E. Kruijff, G. Wesche, K. Riege, G. Goebbels, M. Kunstman, and D. Schmalstieg, "Tactylus, a pen-input device exploring audiotactile sensory binding," in *Proc. ACM Symp. Virtual reality Softw. Technol. (VRST)*, 2006, pp. 312–315.

[38] A. Marquardt, E. Kruijff, C. Trepkowski, J. Maiero, A. Schwandt, A. Hinkenjann, W. Stuerzlinger, and J. Schöning, "Audio-tactile proximity feedback for enhancing 3D manipulation," in *Proc. 24th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2018, pp. 1–10.

[39] M. Samad, E. Gatti, A. Hermes, H. Benko, and C. Parise, "Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–13.

[40] Y. Taima, Y. Ban, T. Narumi, T. Tanikawa, and M. Hirose, "Controlling fatigue while lifting objects using pseudo-haptics in a mixed reality space," in *Proc. IEEE Haptics Symp. (HAPTICS)*, Feb. 2014, pp. 175–180.

[41] D. A. G. Jauregui, F. Argelaguet, A.-H. Olivier, M. Marchal, F. Multon, and A. Lecuyer, "Toward 'pseudo-haptic avatars': Modifying the visual animation of self-avatar can simulate the perception of weight lifting," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 4, pp. 654–661, Apr. 2014.

[42] J. Heron, J. V. M. Hanson, and D. Whitaker, "Effect before cause: Supramodal recalibration of sensorimotor timing," *PLoS ONE*, vol. 4, no. 11, p. e7681, Nov. 2009.

[43] Y. Sugano, M. Keetels, and J. Vroomen, "Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities," *Exp. Brain Res.*, vol. 201, no. 3, pp. 393–399, Oct. 2009.

[44] K. Toida, K. Ueno, and S. Shimada, "Recalibration of subjective simultaneity between self-generated movement and delayed auditory feedback," *NeuroReport*, vol. 25, no. 5, pp. 284–288, Mar. 2014.

[45] V. Jousmäki and R. Hari, "Parchment-skin illusion: Sound-biased touch," *Current Biol.*, vol. 8, no. 6, pp. R190–R191, Mar. 1998.

[46] J. M. Yau, J. B. Olenczak, J. F. Dammann, and S. J. Bensmaia, "Temporal frequency channels are linked across audition and touch," *Current Biol.*, vol. 19, no. 7, pp. 561–566, Apr. 2009.

[47] S. Guest, C. Catmur, D. Lloyd, and C. Spence, "Audiotactile interactions in roughness perception," *Exp. Brain Res.*, vol. 146, no. 2, pp. 161–171, Jul. 2002.

[48] Y. Suzuki, J. Gyoba, and S. Sakamoto, "Selective effects of auditory stimuli on tactile roughness perception," *Brain Res.*, vol. 1242, pp. 87–94, Nov. 2008.

[49] C. Magnusson and K. Rassmus-Gröhn, "A pilot study on audio induced pseudo-haptics," in *Proc. 3rd Int. Haptic Auditory Interact. Design Workshop*, 2008, pp. 6–7.

[50] P. Walker, G. Scallon, and B. Francis, "Cross-sensory correspondences: Heaviness is dark and low-pitched," *Perception*, vol. 46, no. 7, pp. 772–792, Dec. 2016.

[51] I. Herbst and J. Stark, "Comparing force magnitudes by means of vibrotactile, auditory, and visual feedback," in *Proc. IEEE Int. Workshop Haptic Audio Vis. Environ. Appl.*, 2005, pp. 1–5.

[52] E. Sikstrom, A. de Gotzen, and S. Serafin, "Self-charactersics and sound in immersive virtual reality—Estimating avatar weight from footstep sounds," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2015, pp. 283–284.

[53] J. I. D. Campbell and V. A. Thompson, "MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis," *Behav. Res. Methods*, vol. 44, no. 4, pp. 1255–1265, Mar. 2012.

[54] A. M. Okamura, M. R. Cutkosky, and J. T. Dennerlein, "Reality-based models for vibration feedback in virtual environments," *IEEE/ASME Trans. Mechatronics*, vol. 6, no. 3, pp. 245–252, Sep. 2001.

[55] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 2011, pp. 143–146.

[56] O. Schimmel and A. Kohlrausch, "On the influence of interaural differences on temporal perception of noise bursts of different durations," *J. Acoust. Soc. Amer.*, vol. 123, no. 2, pp. 986–997, Feb. 2008.

[57] I. A. Kuling, R. L. J. van Eijk, J. F. Juola, and A. Kohlrausch, "Effects of stimulus duration on audio-visual synchrony perception," *Exp. Brain Res.*, vol. 221, no. 4, pp. 403–412, Sep. 2012.

SEITARO KANEKO (Student Member, IEEE) received the B.E. and M.E. degrees from The University of Electro-Communications, Japan, in 2013 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include tactile sensors and perception. He is a JSPS Research Fellow DC1.

TAKUMI YOKOSAKA received the Ph.D. degree in information processing from the Tokyo Institute of Technology, in 2018. He is employed at NTT Communication Science Laboratories, Japan. His research interests include perception and cognition shaped by body movement.

HIROYUKI KAJIMOTO (Member, IEEE) received the Ph.D. degree in information science and technology from The University of Tokyo, in 2006. He is currently a Professor with the Department of Informatics, The University of Electro-Communications, Japan. His research interests include tactile displays, tactile sensors, human interface, and virtual reality.

TAKAHIRO KAWABE received the Ph.D. degree in psychology from Kyushu University, Fukuoka, in 2005. In 2011, he joined NTT Communication Science Laboratories, where he studies basic and applied aspects of human perception. He is a Senior Distinguished Researcher and the Leader of the Sensory Representation Group, Human Information Science Laboratories, and NTT Communication Science Laboratories. He is a member of the Vision Sciences Society and the Vision Society of Japan. He received the 2013 JPA Award for International Contributions to Psychology: Award for Distinguished Early and Middle Career Contributions from the Japanese Psychological Association. In 2018, he was awarded the Young Scientists' Prize of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology.

• • •