My voice therefore I spoke: sense of agency over speech enhanced in hearing self-voice

Ryu Ohata[1,2,*], Tomohisa Asai[2], Shu Imaizumi[3], and Hiroshi Imamizu[1,2,4*]

[1]Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

[2]Cognitive Mechanisms Laboratories, Advanced Telecommunications Research Institute International (ATR), Keihanna Science City, Kyoto 619-0288, Japan

[3]Institute for Education and Human Development, Ochanomizu University, Otsuka 2-1-1, Bunkyo-ku, Tokyo 112-8610, Japan

[4]Research into Artifacts, Center for Engineering, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

[*] Correspondence and requests for materials should be addressed to H.I (imamizu@gmail.com) or R.O (ryu.oohata@gmail.com).

Abstract

The subjective experience of causing an action is known as the sense of agency. Dysfunction in the sense of agency has been suggested as a cause of auditory hallucinations (AHs), an important diagnostic criterion for schizophrenia. However, agency over speech has not been extensively characterized in previous empirical studies. Here, we examine both implicit and explicit measures of the sense of agency and reveal bottom-up and top-down components that constitute self-agency during speech. The first is action-outcome causality, which is perceived based on a low-level sensorimotor process when hearing their own voice following their speech. The second component is self-voice identity, which is embedded in the acoustic quality of voice and dominantly influences agency over speech at the cognitive judgment level. Our findings provide profound insight into the sense of agency over speech and present an informative perspective for understanding aberrant experience in AHs.

Introduction

The subjective experience that "I" am the one who is causing an action, referred to as the sense of agency, is the fundamental aspect of the sense of self[1, 2]. Although people are not usually aware of its existence, pathological conditions make it evident that this sense is essential in our daily activities. For example, the delusion of control, which is one of the important diagnostic criteria for schizophrenia, denotes an abnormal experience in which an external force controls one's actions, thoughts, or feelings. Theoretical studies hypothesized that these symptoms arise from dysfunction in the sense of agency[3, 4, 5, 6]. Psychological experiments have provided supportive evidence of this hypothesis by examining awareness in patients with schizophrenia during hand/limb movements[7, 8, 9, 10, 11]. Furthermore, it has been suggested that an auditory hallucination (AH), which is also a representative symptom of schizophrenia, also results from dysfunction in the sense of agency over speech[12, 13, 14]. However, compared with the sense of agency over hand/limb movement, the mechanism underlying the agency over speech has not been extensively characterized.

The theoretical model for the sense of agency, known as the comparator model[4, 15], emphasizes the role of the sensorimotor system based on a computational model of motor control. Here, an internal forward model[16] predicts sensory outcomes of action from an efference copy of a motor command and then compares the predicted and actual outcomes in the brain. If the two outcomes are congruent, people perceive causality between their action and consecutive sensory outcomes and, as a result, feel their agency over the action (blue frame in Fig. 1). Neurophysiological studies in humans have captured this comparison process in the sensorimotor system for speech. Auditory event-related potential (N1-component) was more greatly suppressed when an undistorted self-voice was fed back to healthy participants compared to a pitch-distorted voice[17]. This suppression was suggested as the instantiation of the result of the comparison. Importantly, the suppression did not appear in patients with AHs, which implies that dysfunction of the comparator system in speech is associated with AHs[18, 19, 20, 21]. Previous studies have, taken together, demonstrated the importance of the predictive sensorimotor system in

understanding the mechanism of AHs[22, 23]. However, to our knowledge, no study has directly investigated whether such a sensorimotor system for speech enables people to perceive causality between their speech and their own voice, which would be necessary for the sense of agency during speech.

In addition to action-outcome causality, another component that may affect the sense of agency over speech is self-voice identity. In general, people quickly and precisely identify whether a voice is their own or that of another. Thus, a sensory outcome of speech itself contains a sign of the self in its acoustic quality. A theoretical framework suggests that the sense of agency is not simply a direct reflection of a low-level sensorimotor process of comparing an action with its outcome. Context cues, background beliefs, and post-hoc inferences (i.e., rationalizations) also affect agency judgment at the cognitive level[24, 25]. It is therefore plausible that the self-sign embedded in acoustic quality affects the judgment of agency over speech (red frame in Fig. 1). Previous studies have investigated voice attribution (whether a heard voice is attributed to oneself or another) by distorting voice feedback, and they found dysfunction in patients with AHs[26, 27, 28, 29]. Such voice attribution indeed reflects the effect of self-voice identity as well as that of action-outcome causality; however, these two effects have not been properly distinguished. Consequently, relative contribution of each effect (or that of an interaction between them) to the judgment of agency over speech remains poorly understood.

The current study investigated in detail the effects of the action-outcome causality and self-voice identity on the sense of agency over speech using both an implicit measure (without a direct self-report on the agency by participants) and an explicit measure (with a direct self-report). We first examined an action-outcome causality perceived during speech by measuring the temporal compression of a perceived interval between speech and voice feedback. This compression, termed the intentional binding effect, is often used as an implicit measure of the sense of agency[2, 30, 31, 32, 33, 34, 35]. The sensory processing for detecting action (i.e., speech) and outcome (i.e., feedback voice) timings is considered independently of the judgment of whether the feedback voice is one's own (i.e., self-voice identity

judgment). Thus, the implicit measure can access a sensorimotor process at a lower level than the cognitive process of self-voice identity. We distorted the pitch of a feedback voice and examined whether the perceived interval would be more compressed in the pitch-undistorted than in the pitch-distorted condition. Our second experiment examined the effect of self-voice identity on the judgment of agency. We required participants to explicitly report on their agency (how much they felt they had caused the voice to be heard) when we manipulated the self-voice identity by giving a pitch-distorted or undistorted self-voice following their speech as voice feedback. A time interval was inserted between their speech and the voice feedback. According to previous studies on hand/limb movement, an action-outcome temporal mismatch is critical to the loss of causality between them, which eventually causes loss of the sense of agency[36, 37, 38, 39, 40, 41, 42]. Thus, we manipulated the interval length to examine how self-voice identity interacts with causality in the judgment of agency. Finally, we examined an association of proneness to AH with implicit and explicit measures of agency.

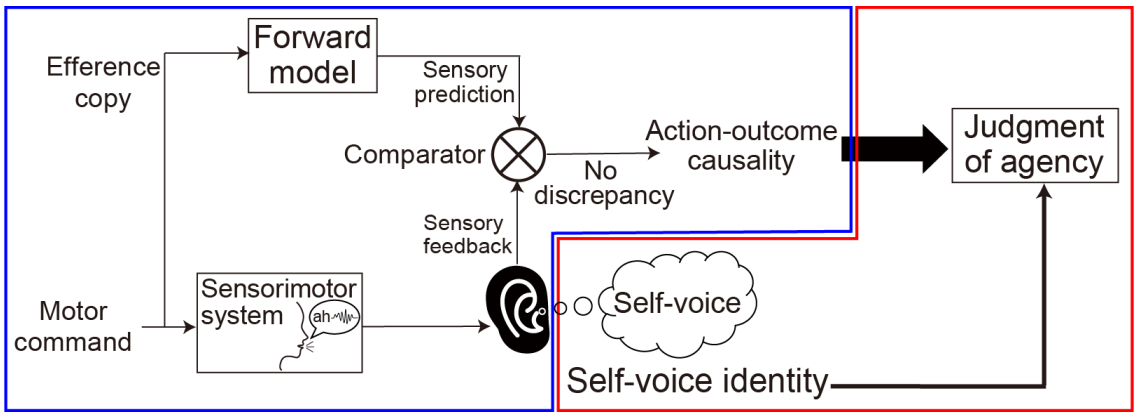

Fig. 1 Hypothetical processes behind the sense of agency over speech. First, a sensory outcome of speech is predicted by the internal forward model. The sensory prediction is compared with actual sensory feedback. The discrepancy between them determines the estimation of causality between action and outcome (i.e., between speech and feedback voice). Sensorimotor processes are highlighted in the blue frame. By contrast, the self-voice identity (whether the feedback voice is one's own) affects the judgment of agency over speech. The cognitive process shown in the red frame also constitutes the sense of agency.

Results

Experiment 1: intentional binding during speech

In the first experiment, we examined the sense of agency over speech in the intentional binding paradigm (Fig. 2a; for details, see Methods: Experiment 1). Twenty-nine participants joined the experiment. They vocalized a vowel sound from the Japanese syllabary ("ah," "i," "u," "e," or "o") into a microphone. Following a 200-ms, 400-ms, or 600-ms interval, they heard their voice through a headphone. The feedback voice was pitch-shifted upward or downward by seven semitones or presented without any distortion (high-/low-pitch and neutral conditions, respectively; see Supplementary Movie for a voice sample under each condition). Participants then reported an estimated interval between their speech and the voice feedback using a numeric keypad. We randomly presented the 45 conditions (three levels of speech-feedback interval × three levels of voice distortion × five levels of syllable sound).

We analyzed estimated intervals using a 3 × 3 repeated-measures analysis of variance (ANOVA) with within-subject factors of pitch and speech-feedback interval (Fig. 2b and "Data Summary" in Supplementary Data). We found a significant main effect of pitch ($F(2, 56) = 18.3$, $p < 0.001$, $\eta_p^2 = 0.39$) and a significant main effect of speech-feedback interval ($F(1.22, 34.3) = 245.9$, $p < 0.001$, $\eta_p^2 = 0.90$, Greenhouse-Geisser corrected) but no significant interaction ($F(2.74, 76.68) = 0.76$, $p = 0.51$, $\eta_p^2 = 0.026$, Greenhouse-Geisser corrected). Following the previous study on intentional binding[43], we calculated the mean of the estimated intervals across the three speech-feedback intervals to examine a simple effect of pitch (Fig. 2c; see also Supplementary Fig. 1 for each speech-feedback interval). We compared the means in the neutral condition with those in the high- and low-pitch conditions. The comparisons revealed that the estimated intervals in the neutral condition were shorter than those in the high-pitch ($t(28) = 5.54$, $p < 0.001$, Cohen's $d = 1.03$, two-tailed paired t-test with Bonferroni correction) and low-pitch conditions ($t(28) = 4.90$, $p < 0.001$, Cohen's $d = 0.91$, two-tailed paired t-test with Bonferroni correction). Collectively, the distorted voice feedback weakens the

compression of the perceived interval between speech and voice feedback.

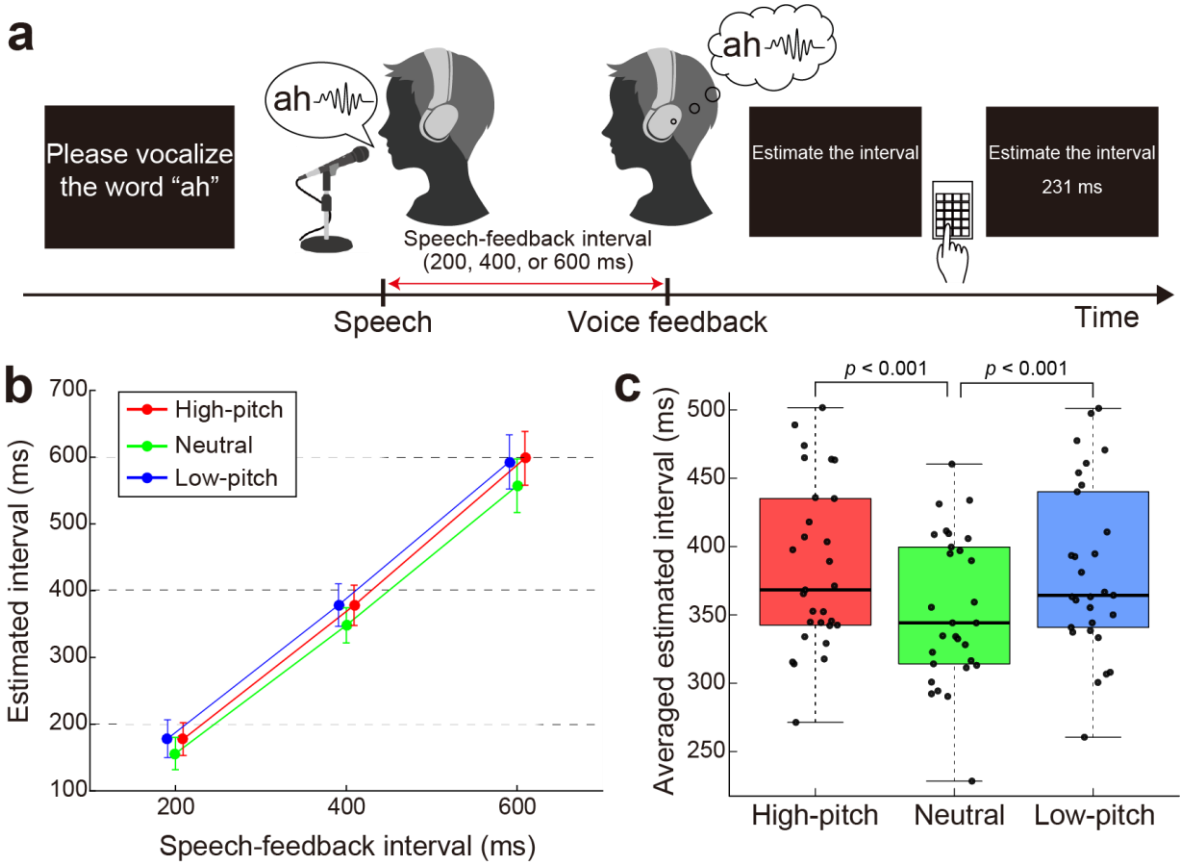

Fig. 2 Experiment 1. (**a**) Intentional binding paradigm during speech. Participants vocalized a sound and heard their voice through a headphone following a short interval (200, 400, or 600 ms). They reported a perceived interval between their speech and voice feedback. (**b**) Mean of the estimated intervals in each pitch and speech-feedback interval. Red, green, and blue circles and lines correspond to the high-pitch, neutral, and low-pitch conditions, respectively. Error bars indicate 95% confidence intervals (CIs). Horizontal dotted lines denote the actual speech-feedback interval. Note that the circles of the high- and low-pitch conditions are shifted slightly rightward and leftward, respectively, for display purpose. (**c**) Means of the estimated intervals across the three speech-feedback intervals. In each box plot, the central horizontal line indicates the median, and the bottom and top edges correspond to the 25th and 75th percentiles. Each dot corresponds to one participant.

Control experiment for Experiment 1

We were concerned that the participants would feel difficulty in detecting the onset of the distorted voice, which might cause longer estimated intervals in the high- and low-pitch conditions than those in the neutral condition. To examine this possibility, we conducted a control experiment in which the same participants in Experiment 1 reported estimated intervals between the perceived time of a beep and a prerecorded voice without online vocalization (Fig. 3a). We analyzed the estimated intervals using a 3 × 3 repeated-measures ANOVA with within-subject factors of pitch and beep-voice interval (Fig. 3b and "Data Summary" in Supplementary Data). As a result, we found a significant main effect of beep-voice interval ($F(1.28, 35.97) = 434.2$, $p < 0.001$, $\eta_p^2 = 0.94$, Greenhouse-Geisser corrected), but no main effect of pitch was significant ($F(2, 56) = 1.04$, $p = 0.36$, $\eta_p^2 = 0.036$). An interaction of these two factors was significant ($F(4, 112) = 3.13$, $p = 0.018$, $\eta_p^2 = 0.10$). We found that the estimated interval at the 200-ms beep-voice interval in the high-pitch condition was significantly shorter than that in the neutral condition by a post hoc paired *t*-test with Bonferroni correction ($t(28) = 3.23$, $p = 0.0096$). However, this effect was not associated with the main result in Experiment 1 (i.e., the shorter estimated interval in the neutral condition; Fig. 2b). As with the analysis of Experiment 1, we compared the mean of the estimated intervals across the three beep-voice intervals in the neutral condition with that in the high-pitch and low-pitch conditions (Fig. 3c; see also Supplementary Figs. S2a, S2b, and S2c for each beep-voice interval). There was no significant difference in the mean between the high-pitch and neutral conditions ($t(28) = 0.72$, $p = 0.48$, Cohen's $d = 0.13$, two-tailed paired *t*-test) or between the low-pitch and neutral conditions ($t(28) = 0.62$, $p = 0.54$, Cohen's $d = 0.12$, two-tailed paired *t*-test).

In addition, previous experimental and computational model studies have shown that the reliability of an outcome sensory signal affects the intentional binding effect[44, 45]. To assess the reliability of auditory perception in the different pitch conditions, we calculated the coefficient of variation (CV; ratio of standard deviation to the mean) of the estimated intervals (Fig. 3d and "Data Summary" in Supplementary Data). A 3 × 3 repeated measures ANOVA revealed a main effect of interval was

significant ($F(1.49, 41.8) = 52.5$, $p < 0.001$, $\eta_p^2 = 0.65$, Greenhouse-Geisser corrected). However, neither a main effect of pitch ($F(1.5, 42.12) = 0.50$, $p = 0.56$, $\eta_p^2 = 0.017$, Greenhouse-Geisser corrected) nor an interaction ($F(1.85, 51.94) = 1.34$, $p = 0.27$, $\eta_p^2 = 0.046$, Greenhouse-Geisser corrected) was significant. Moreover, we found no effect of the pitch by averaging the CVs across the three beep-voice intervals (Fig. 3e; see also Supplementary Figs. S2d, S2e, and S2f for each beep-voice interval). Taken together, neither difficulty in detecting onset nor the reliability of auditory perception influenced the results in Experiment 1. Hence, the more compression found in the neutral condition than in the pitch-distorted conditions likely reflected the difference in the sense of agency. Also, although the pitch distortion might compress or expand the length of voice sound, this control experiment indicates that the changes in the length, if any, unlikely affected the difference in estimated intervals among the conditions.

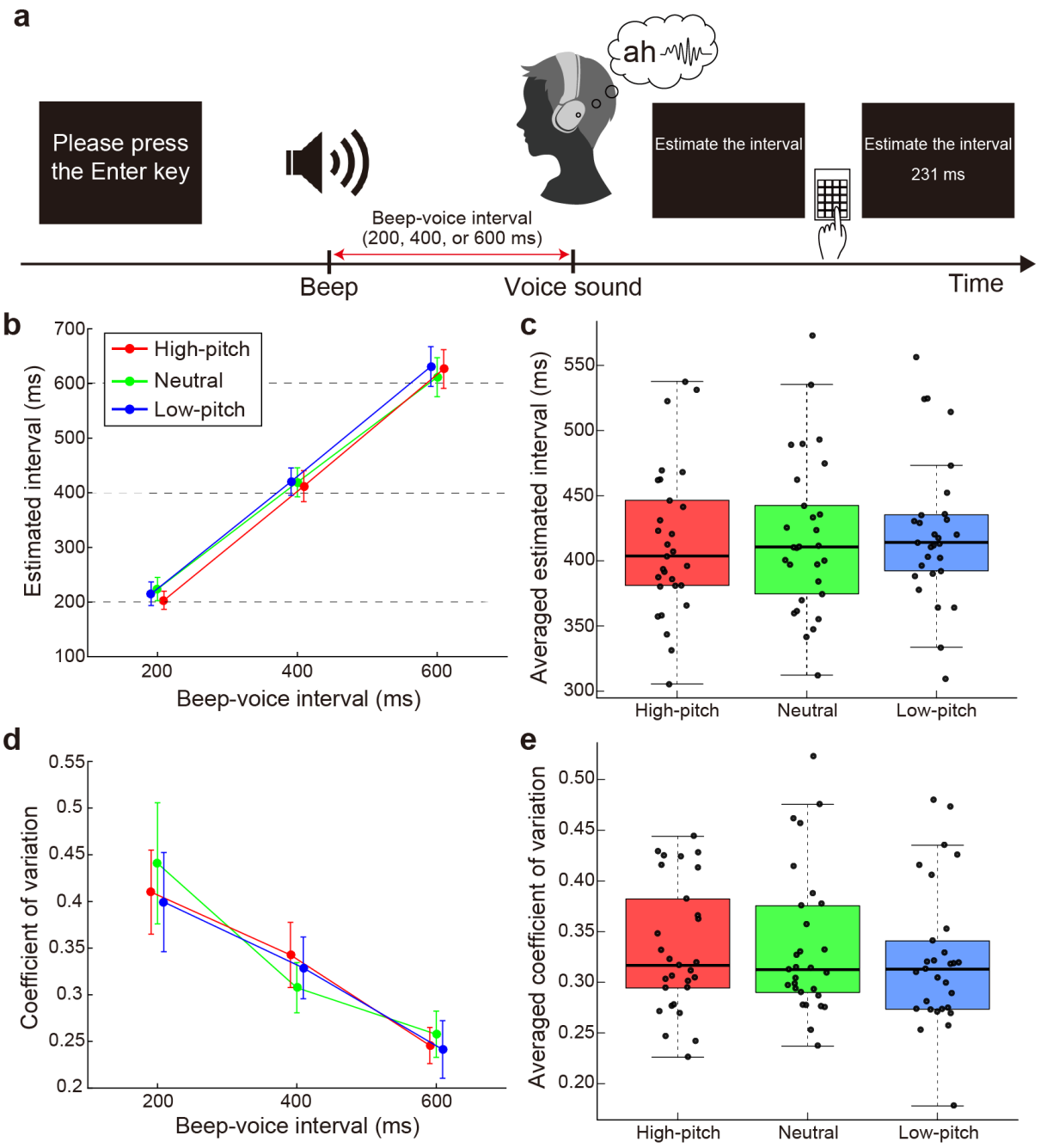

Fig. 3 Control experiment for Experiment 1. (**a**) Experimental task. Participants heard their prerecorded voice 200, 400, or 600 ms after a beep. They reported a perceived interval between the beep and voice sound. (**b**-**e**) The mean (**b**) and coefficient of variation (**d**) of estimated intervals are shown for each pitch and beep-voice interval condition. Red, green, and blue circles and lines correspond to the high-pitch, neutral, and low-pitch conditions, respectively. Error bars indicate 95% CIs. Horizontal dotted lines in (**b**) denote the actual beep-voice interval. Note that the circles of high- and low-pitch conditions are shifted slightly rightward and leftward, respectively, for display purpose. The estimated intervals (**c**) and coefficient of variation (**e**) are averaged across the three beep-voice intervals. In each box plot, the central horizontal line indicates the median, and the bottom and top edges correspond to the 25th and 75th percentiles. Each dot corresponds to one participant.

Experiment 2: subjective ratings on agency and self-voice identity

In the second experiment, we required participants to give explicit reports regarding agency over speech and self-voice evaluation (Fig. 4a; for details, see Methods: Experiment 2). Twenty-eight participants who participated in Experiment 1 joined this second experiment. As with the procedure in Experiment 1, they vocalized a vowel sound in the Japanese syllabary and heard their voice through a headphone following a short interval (50, 200, 350, 500, or 650 ms). They then reported a subjective rating regarding agency (agency rating: "how much did you feel you caused the voice to be heard?") or self-voice evaluation (self-voice rating: "how much did you feel the voice heard was your own?") on a 9-point Likert scale. The feedback-voice was pitch-shifted upward or downward by seven semitones or presented without distortion (high-/low-pitch and neutral conditions, respectively).

Figure 4b shows the mean score of agency rating in each pitch and speech-feedback interval (see also "Data Summary" in Supplementary Data). We conducted a 3 × 5 repeated-measures ANOVA with within-subject factors of pitch and speech-feedback interval. As a result, a main effect of pitch was significant ($F(1.39, 37.49) = 79.3$, $p < 0.001$, $\eta_p^2 = 0.75$, Greenhouse-Geisser corrected), but no main effect of speech-feedback interval was significant ($F(1.15, 30.96) = 2.67$, $p = 0.11$, $\eta_p^2 = 0.090$, Greenhouse-Geisser corrected). In addition, an interaction between these two factors was significant ($F(4.85, 130.98) = 5.10$, $p < 0.001$, $\eta_p^2 = 0.16$, Greenhouse-Geisser corrected). Post-hoc tests revealed a significant simple main effect of speech-feedback interval in the high-pitch condition ($F(1.32, 35.6) = 4.29$, $p = 0.036$, $\eta_p^2 = 0.14$, Greenhouse-Geisser corrected) and a marginal effect in the low-pitch condition ($F(1.29, 34.95) = 3.27$, $p = 0.069$, $\eta_p^2 = 0.11$, Greenhouse-Geisser corrected), but no significant effect was found in the neutral condition ($F(1.28, 34.52) = 0.80$, $p = 0.40$, $\eta_p^2 = 0.029$, Greenhouse-Geisser corrected).

To our surprise, we did not find a significant effect of the speech-feedback interval, at least for the neutral condition, despite the fact that a temporal mismatch between an action and its outcome is critical for a decrease in the sense of agency in hand/limb movement tasks[36, 37, 38, 39, 40, 41, 42]. Next, we

compared the results using our speech paradigm with those using the button-press paradigm reported by Imaizumi and Tanno (2019)[39]. In this previous study (Fig. 4d), participants pressed a button followed by a short interval (100, 300, 500, 700, or 900 ms) and heard a tone via a headphone. The participants then rated the perceived sense of agency over the tone by answering the question: "How much did you feel that you had caused the tone?" We fitted a linear regression model to each of 34 participants' rating score as a function of action-outcome interval (ms). We also fitted the model to the agency ratings in Experiment 2. As a result, the mean of the slopes was -5.6 (SD: 2.1) in the button-press task, whereas the means were -0.47 (3.6), -1.7 (3.9), and -1.6 (4.1) in the neutral, high-pitch, and low-pitch conditions of the speech task, respectively (Fig. 4e). The slope in the button-press task was significantly and negatively steeper than the values under any condition of the speech task (high-pitch condition: $t(60) = 4.99$, $p < 0.01$, Cohen's $d = 5.0$; neutral condition: $t(60) = 7.00$, $p < 0.01$, Cohen's $d = 5.45$; and low-pitch condition: $t(60) = 4.99$, $p < 0.01$, Cohen's $d = 5.12$, two-tailed t-test). Furthermore, we compared the results using our speech paradigm with those from the two different tasks using hand movement[36, 38]. Note that participants in these studies reported agency judgment differently from those in Imaizumi and Tanno (2019) (i.e., 9-point Likert scale versus two-alternative forced-choice). Thus, the procedure in our study was more similar to that in Imaizumi and Tanno (2019) than those in the two hand-movement tasks. We found the slopes in the hand-movement tasks were significantly and negatively steeper than the values in at least the neutral condition of the speech task (for details, see Supplementary Texts: 1. Comparison of agency judgment between hand/limb movement and speech tasks and Supplementary Fig. 3). These comparisons clarified that the judgment of agency over speech was more immune to action-outcome mismatch than that of agency over hand/limb movement.

Next, regarding self-voice rating (Fig. 4c and "Data Summary" in Supplementary Data), we found a significant main effect of pitch ($F(1.56, 42.16) = 274.9$, $p < 0.001$, $\eta_p^2 = 0.91$) but neither a significant main effect of speech-feedback interval ($F(1.5, 40.45) = 0.94$, $p = 0.37$, $\eta_p^2 = 0.034$) nor a significant interaction ($F(4.73, 127.58) = 0.19$, $p = 0.96$, $\eta_p^2 = 0.0070$). Thus, both agency rating and

self-voice rating were sensitive to the pitch distortion of the feedback voice, but the self-voice rating was not significantly affected by the speech-feedback interval.

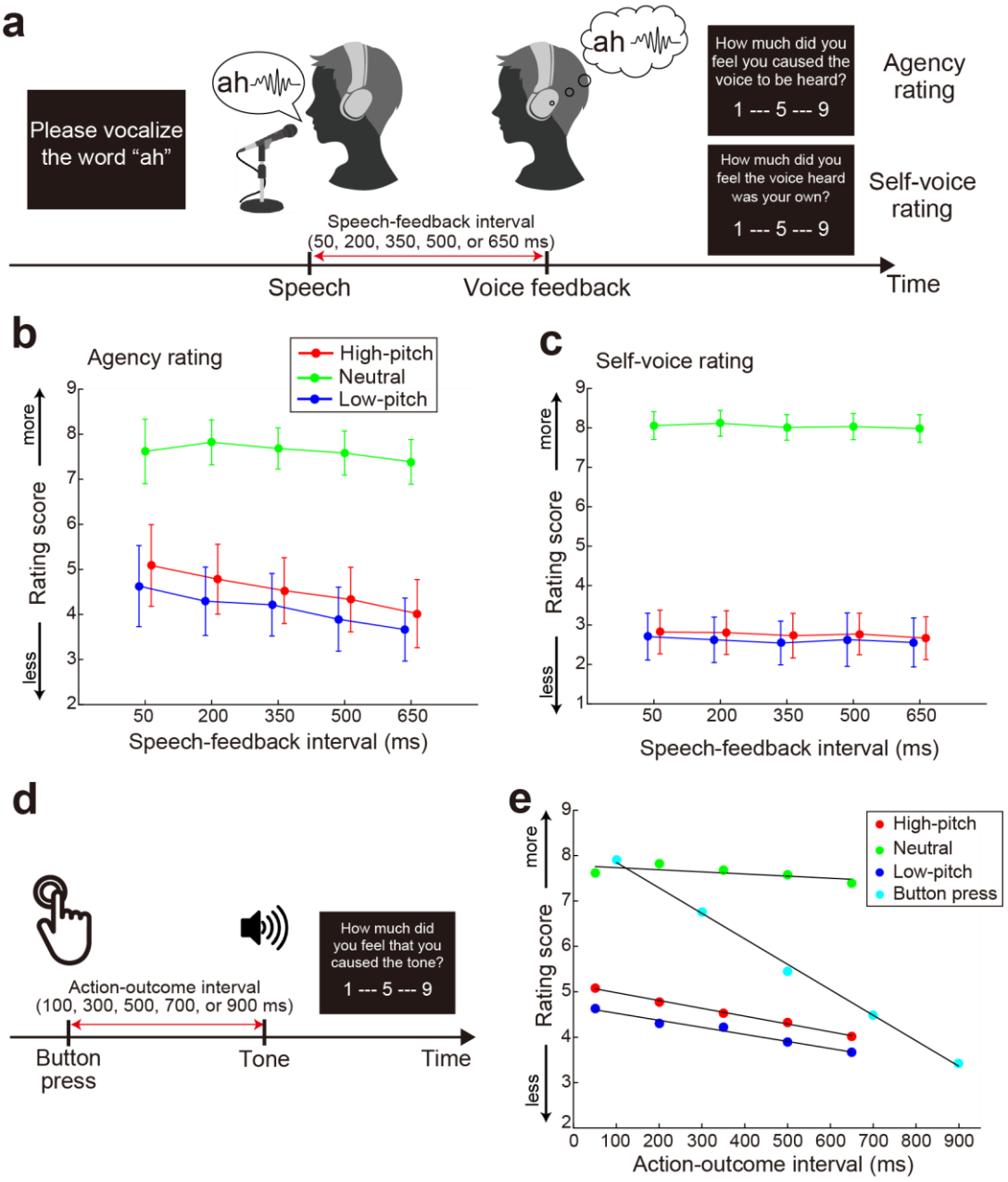

Fig. 4 Experiment 2. (**a**) Experimental task for explicit reporting on agency and self-voice evaluation. Participants vocalized a sound and heard a feedback voice after a short interval (50, 200, 350, 500, or 650 ms). They reported how much they felt that they had caused the heard voice (agency rating) or how much they felt the heard voice was their own (self-voice rating) on a 9-point Likert scale. (**b** and **c**) Means of agency rating (**b**) and self-voice rating (**c**) in each pitch and speech-feedback interval. Red, green, and blue circles and lines correspond to high-pitch, neutral, and low-pitch conditions, respectively. Error bars indicate 95% CIs. Note that the circles of high- and low-pitch conditions are shifted slightly rightward and leftward, respectively, for display purpose. (**d**) Button-press task in Imaizumi & Tanno (2019). Participants pressed a button and heard a tone sound after a short interval (100, 300, 500, 700, or 900 ms). They reported how much they felt that they had caused the tone on a 9-point Likert scale. (**e**) Comparison of agency ratings in the speech task of the current study with those

in the button-press task. The means of agency ratings are plotted as a function of action-outcome interval. Red, green, and blue circles correspond to high-pitch, neutral, and low-pitch conditions, respectively. Cyan circle denotes the mean of agency ratings in the button-press task. Black lines indicate regression lines, which were fitted to the means of ratings in the respective conditions.

Association with proneness to having auditory hallucinations

Finally, we examined whether the indices measured in Experiments 1 and 2 (estimated intervals, agency ratings, and self-voice ratings in the three pitch-conditions) could explain the individual difference in proneness to having AHs, as quantified by a questionnaire (Auditory Hallucination Experience Scale 17: AHES-17). We conducted a stepwise multiple regression analysis, using MATLAB function *stepwisefit*, to add or remove indices according to how their inclusion affects the model. The indices that significantly improved the model (F statistic, $p < 0.05$) were added to the model as a predictor. As a result, the self-voice ratings in the low-pitch condition alone could significantly explain the AHES-17 scores ($R^2 = 0.24$, $F(1,26) = 8.24$, $p = 0.0080$). Figure 5 shows the correlation between the self-voice ratings in the low-pitch condition and AHES-17 scores. The correlation coefficient was significantly larger than zero according to a permutation test ($r = 0.49$, $p = 0.0051$; 10,000 times randomization). Thus, the above results indicate that individual differences in proneness to AH were associated with the process of evaluating self-identity in hearing a voice.

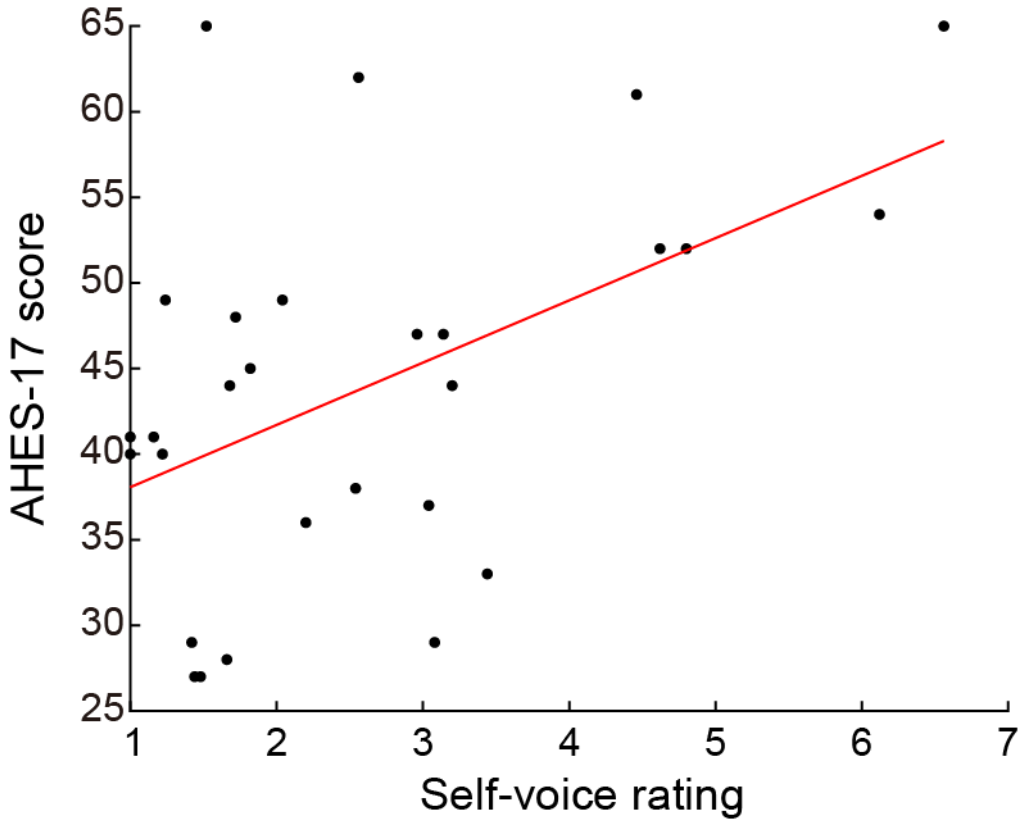

Fig. 5 Correlation between self-voice rating in low-pitch condition and proneness to auditory hallucinations (AHES-17 score). Correlation coefficient was 0.49 (p = 0.0051, permutation test). Each dot corresponds to one participant. The red line denotes the regression line.

Discussion

The current study examined the two possible components that affect the sense of agency over speech: action-outcome causality and self-voice identity. In the first experiment, we found more compression of the estimated intervals between speech and voice feedback when the participants heard an undistorted self-voice than a distorted one (Figs. 2b and 2c). This compression, corresponding to an intentional binding effect, indicates that stronger action-outcome causality can be perceived in hearing self-voice following one's speech. In the second experiment, we investigated the effect of self-voice identity, manipulated by pitch distortion, on the judgment (rating) of agency. We first found that self-voice identity strongly affects the agency rating (Fig. 4b) and the self-voice rating (Fig. 4c). We next found that the agency rating over speech was immune to a temporal mismatch between action and its outcome (i.e., between speech and feedback voice), especially in the case where self-voice identity was preserved (i.e., when undistorted self-voice was fed back) (Fig. 4e). This property was not confirmed in the judgment of agency over hand/limb movement in previous studies. Finally, we found an association of proneness to AH with the self-voice rating but not with either an implicit or explicit measure of agency (Fig. 5).

In Experiment 1, we demonstrated, for the first time to our knowledge, the intentional binding effect in a naturalistic speech paradigm (but see Limerick et al. (2015)[46], which used a speech interface controlled by voice command). What caused the lesser compression in the pitch-distorted conditions than in the neutral condition? A recent Bayesian model of the sense of agency rationally explains the intentional binding effect. It posits three requirements for confidence in causal estimation, which is how the model theorizes the sense of agency[44]: (1) temporal consistency in the action-outcome interval, (2) prior belief that the action caused the outcome, and (3) reliability of sensorimotor signals informing action and outcome timings. In the current experiment, we randomly presented three speech-feedback intervals and three pitch-distorted conditions on a trial-by-trial basis. Therefore, regarding the first two of the above requirements, temporal consistency in the action-outcome effect and prior causal belief are

comparable in all pitch-distorted conditions. Thus, referring to the computational model, a plausible cause of the lesser compression in the pitch-distorted conditions is having unreliable sensorimotor signals of action and outcome timings. As suggested by the comparator model[4, 15], the sensorimotor system predicts a sensory outcome of speech (i.e., feedback voice) and then compares the predicted and actual outcomes[13, 22, 23]. People already establish an internal forward model of speech in the course of daily activities. As the established internal model predicts non-distorted voice as an outcome of speech, the pitch distortion produces a discrepancy in acoustic quality between the predicted and actual outcomes. The neural activity involved in the sensorimotor processing may be perturbed by this discrepancy, thus reducing the sensory reliability of action and outcome timings and estimating weakened causality between speech and voice feedback. This explanation is consistent with the perturbed neural activity in hearing distorted voice feedback found in previous neurophysiological studies[13, 17, 20]. Accordingly, such a low-level sensorimotor process plausibly lessened the compression in the pitch-distorted conditions.

Experiment 2 investigated the effect of self-voice identity on the sense of agency over speech at the cognitive judgment level. The first finding in this experiment was the large effect of self-voice identity on agency over speech. The pitch distortion remarkably decreased the agency rating scores (Fig. 4b) as well as the self-voice rating (Fig. 4c), and these decreases were almost comparable. This effect appeared in the vertical discrepancies between the rating scores in the neutral condition and those in the two distorted conditions in Figs. 4b and 4c. These results suggest that the sign of the self in the acoustic quality of feedback voice strongly affects the judgment of agency over speech.

The second finding in Experiment 2 was a small effect of speech-feedback interval on the agency rating. A temporal mismatch between an action and its outcome is generally a critical factor in reducing the sense of agency[36, 37, 38, 39, 40, 42]. However, compared with the button-press paradigm[39], we found that the agency ratings less steeply decreased as a function of action-outcome interval in all pitch conditions of this speech paradigm (Fig. 4e). The critical difference between the button-press paradigm

and our speech paradigm is whether the outcome of an action (i.e., tone or voice sound, respectively) contains a sign of the self. In our additional experiment (see Supplementary Texts: 2. Self-voice identification of prerecorded voice as a function of pitch distortion), we confirmed that 1) the participants recognized the undistorted voice as their own and 2) the level of pitch distortion in our experiment (i.e., ±7 semitones) was adequate for them to recognize the distorted voice as a non-self-voice. These results suggest that participants linked the feedback voice with their speech once they recognized the voice as their own, whereas this link was disconnected once they recognized the voice as someone else's. This top-down cognitive effect of self-voice identity was so dominant that the agency judgment became immune to action-outcome temporal relation that could contribute to the agency judgment as a bottom-up process. Thus, our findings highlighted the dominant top-down effect of self-voice identity on the judgment of agency over speech (red frame in Fig. 1).

The third finding in Experiment 2 was the significant interaction between pitch and action-outcome interval shown by the ANOVA of the agency rating scores. The post-hoc tests for individual pitch-conditions revealed that the simple main effect of the interval was significant only in the pitch-shifted conditions (red and blue lines in Fig. 4b), but not in the neutral condition (green line in Fig. 4b). We can explain the reasons for this interaction in terms of the dominance of self-voice identity as follows. On the one hand, a non-distorted self-voice was heard following a speech in the neutral condition. The participants might perceive this phenomenon as reasonable based on their daily experience. Thus, self-voice identity dominantly affected the agency rating in this condition, resulting in the rating being immune to action-outcome temporal mismatch. On the other hand, it is unreasonable to hear a distorted self-voice following a speech presented in the pitch-distortion conditions. Thus, the dominance of the self-voice identity was relatively weaker in those conditions than in the neutral condition. Therefore, a mixed effect of self-voice identity and action-outcome temporal mismatch determined the agency ratings.

In addition to the above findings, we found that neither implicit nor explicit measures of the

sense of agency but instead self-voice rating in the low-pitch condition could explain the individual differences in proneness to AHs (Fig. 5). In addition to the low-pitch condition, we found a moderate correlation between the self-voice ratings in the high-pitch condition and AHES-17 scores (r = 0.32, p = 0.045, uncorrected, permutation test with 10,000 times randomization). By contrast, the correlations with the self-voice ratings in the neutral condition and with the other indices (i.e., estimated intervals and agency ratings) scored at most r = 0.10. These results suggest that the cognitive process of evaluating self-voice identity in a feedback voice is critical to explaining the auditory hallucination-like experiences in (non-psychotic) participants. One possible reason why only the self-voice ratings were associated with proneness to AH is because the scale we evaluated in this study was related to proneness not to schizophrenic but to dissociative AHs. Although AHs are indeed a representative symptom of schizophrenia, clinical observations have shown that patients with dissociative identity disorder also exhibit first-rank symptoms of schizophrenia, including AHs[47, 48]. Unlike pathological dissociation, even some healthy individuals can experience mild dissociation, such as talking with imaginary friends inside the head. Such mild dissociation might moderately disrupt self-voice identity. Thus, if the AHES-17 scores mainly reflect proneness to dissociative (not schizophrenic) AH in the current study, it is reasonable that, as found in our results (Fig. 5), these scores could be correlated with people's ability to precisely evaluate the self-voice identity of a feedback voice. To validate this possibility, we need further studies that compare the abilities of participants showing schizotypal traits with those showing dissociative traits.

In the present study, there is a limitation on bone conduction's influence upon the sense of agency over speech. Voices heard through our ears are not the only sensory feedback in speech. Bone conduction is also one of the primary sensory inputs active while we are speaking. We cannot deny bone conduction's effect on the sense of agency measured in the present task, although this effect was reduced by mixing white noise into the feedback voice (see Methods: Apparatus). However, the effect should not be different across the pitch conditions. Therefore, bone conduction unlikely affected our main

conclusion that distorted self-voice feedback attenuated the sense of agency over speech.

In sum, this study has characterized the mechanism underlying the sense of agency over speech. The first experiment indicated that the predictive sensorimotor system enabled the brain to estimate causality between an action and its outcome during speech (blue frame in Fig. 1). This low-level bottom-up process has also been characterized as the primary component of the sense of agency over hand/limb movement. The second experiment, by contrast, highlighted the uniqueness of the sense of agency over speech; that is, the sign of the self in the acoustic quality of voice (i.e., self-voice identity) constitutes a crucial component of the sense of agency at the cognitive judgment level (red frame in Fig. 1). This top-down effect was so dominant that the judgment is immune to a result processed by the low-level sensorimotor system. These findings shed light on the nature of the subjective experience of self-agency during speech. They would also be informative for a deeper understanding of aberrant experience in AHs.

Methods

Participants

Twenty-nine healthy volunteers (15 males and 14 females) with a mean age of 21.7 (19–25 years of age) participated in Experiment 1 and the control experiment for Experiment 1. The same volunteers, except for one female (i.e., 28 volunteers), participated in Experiment 2. We determined the sample sizes based on our preliminary experiment (see Supplementary Texts: 3. Preliminary experiment on the effect of intentional binding during speech). For the sample size calculation, we performed a power analysis for repeated measures ANOVA using G*power 3.1 with power selected at 0.95, effect size (η_p^2) at 0.42, and alpha at 0.05[49]. The experimental protocol was approved by the ethics committee at the University of Tokyo. Written informed consent was obtained from all volunteers in accordance with the latest version of the Declaration of Helsinki.

Apparatus

All experiments were conducted inside a soundproof room. Participants seated themselves in front of a microphone (SENNHEISER MD42) on a desk and wore a headphone (HyperX Cloud Revolver Pro Gaming Headset, Kingston Technology Company). The microphone was connected to an effector (BEHRINGER VIRTUALIZER Pro) via an amplifier (AT-MA2 MICROPHONE AMPLIFIER, Audio-Technica). We used the effector to change the pitch of a spoken voice and to insert a time interval between speech and voice feedback. Note that the effector's sample rate was so high (46 kHz) that people could not detect any lag due to the experiment's pitch distortion. White noise was mixed into the feedback voice using a sound mixer (YAMAHA MW10C) to reduce direct voice transmission (not via the effector and headphone) and bone conduction. A custom Python program running on a laptop computer placed on a desk presented the visual stimuli and collected the participants' responses.

Questionnaire on proneness to auditory hallucination

Before the participants performed the task of Experiment 1, they answered a questionnaire on their proneness to auditory hallucination (AH). We used the Auditory Hallucination Experience Scale 17 (AHES-17), which was translated into Japanese[50] and had been used in previous studies[51, 52, 53]. The AHES-17 is a short version of the AHES and includes 17 self-reporting questions, each of which participants answer on a 5-point Likert scale. The range of possible scores is from 17 to 85.

Experiment 1: Intentional binding during speech

***Experimental task and procedure*:** Participants were required to report the perceived interval between their speech and voice feedback (Fig. 2a). At the beginning of each trial, a message was presented on the screen to instruct them to utter one of five vowel sounds in the Japanese syllabary ("ah," "i," "u," "e," or "o"). Participants pressed the enter key when they were ready, and the message on the screen disappeared. Then, they vocalized the instructed sound into the microphone. Following a short interval (200, 400, or 600 ms), participants heard the spoken sound through a headphone. The feedback voice was pitch-shifted upward or downward by seven semitones or presented without any distortion (high- or low-pitch or neutral condition, respectively). After the message "Estimate the interval" was presented on the screen, they reported the estimated interval between their speech and voice feedback using a numeric keypad. They were instructed to vocalize a sound as briefly and clearly as possible, and they were required to speak closely to the microphone. Participants reported the estimated interval in three digits, ranging from 100 to 999 ms. They performed four trials for each of the 45 conditions (three levels of speech-feedback interval × three levels of voice distortion × five levels of sound) in random order (i.e., 180 trials in total).

Before the main task session, participants performed a practice session to get accustomed to the estimation of an interval between speech and voice feedback. They vocalized the sound instructed on the screen and heard their spoken voice, as in the main task. In the practice session, we set a speech-feedback interval at 11 different levels from 0 to 1000 ms with a 100-ms interval. The feedback voice was not distorted (i.e., the neutral condition only). Participants conducted 22 trials in total (2 trials × 11

conditions in random order).

***Data analysis*:** We averaged estimated intervals across the five sounds in each pitch and speech-feedback interval condition for each participant. Then, we performed a 3 × 3 repeated-measures ANOVA on the estimated intervals with within-subject factors of pitch (high, neutral, and low) and speech-feedback interval (200, 400, and 600 ms). If the sphericity assumption for ANOVA was violated, we applied Greenhouse–Geisser correction.

Control experiment for Experiment 1

***Experimental task and procedure*:** Participants were required to report the interval between the perceived time of a beep and a prerecorded voice sound (Fig. 3a). At the beginning of a trial, a message was presented on the screen prompting participants to press the enter key. A beep sound was presented after a random interval (2,000 to 5,000 ms) from the keypress. Following a short interval (200, 400, or 600 ms), a recorded voice was replayed through a headphone. Then, the participants reported the estimated interval between the time of a beep and voice sound using a numeric keypad. They were instructed to report the estimated interval in three digits, ranging from 100 to 999 ms. Before the experimental task, we recorded each participant's voice as they vocalized five vowel sounds in the Japanese syllabary ("ah," "i," "u," "e," or "o"). We extracted vocalization sections from the recorded auditory file. We distorted the pitch by shifting it upward and downward by seven semitones (high- and low-pitch condition, respectively) using audio software (Audacity, https://www.audacityteam.org/). Participants conducted four trials for each of the 45 conditions (three levels of beep-voice interval × three levels of voice distortion × five levels of sound) in random order (i.e., 180 trials in total). They performed a practice session before the main task session. In the practice session, the beep-voice interval was set from 0 to 1000 ms with 100-ms intervals. The recorded voice was replayed without pitch distortion. Participants performed two trials for each of the 11 conditions in random order (i.e., 22 trials in total).

***Data analysis*:** We averaged estimated intervals across the five sounds in each pitch and beep-voice

interval condition for each participant. We performed a 3 × 3 repeated-measures ANOVA on estimated intervals with within-subject factors of pitch (high, neutral, and low) and beep-voice interval (200, 400, and 600 ms). If the sphericity assumption for ANOVA was violated, we applied Greenhouse–Geisser correction. In the post-hoc test for the ANOVA, Bonferroni correction was used for multiple comparisons.

Experiment 2: Subjective report on agency over speech and self-voice evaluation

***Experimental task and procedure*:** Participants were required to report subjective ratings regarding agency over speech and self-voice identification (Fig. 4a). At the beginning of each trial, a message was presented on the screen to instruct them to vocalize one out of five sounds in the Japanese syllabary ("ah," "i," "u," "e," or "o"). The message disappeared immediately after the keypress, and participants spoke the instructed sound into the microphone. Following a short interval (50, 200, 350, 500, or 650 ms), the spoken sound was heard through the headphone. The feedback-voice was pitch-shifted upward or downward by seven semitones or presented without distortion (high- or low-pitch or neutral condition, respectively). After a message was presented on the screen, they reported a subjective rating regarding agency over speech (agency rating) or regarding self-voice evaluation (self-voice rating). The message was "how much did you feel you had caused the voice to be heard?" for the agency rating or "how much did you feel the voice heard was your own?" for the self-voice rating. The rating was scored on a 9-point Likert scale from 1 (not at all) to 9 (definitely). As in Experiment 1, we instructed participants to vocalize the sound as briefly and clearly as possible and to speak closely to the microphone.

In Experiment 2, we did not require participants to perform any practice session. We were concerned about the possibility that they might confuse the agency rating with the self-voice rating if they reported both in a single trial or either in random order. Therefore, we required them to report agency ratings in the first half block and then self-voice ratings in the last half block. In each block, they conducted two trials for each of 75 conditions (five levels of speech-feedback interval × three levels of voice distortion × five levels of sound) in random order. Together with the two blocks, they

conducted 300 trials in total.

***Data analysis*:** We averaged rating scores across the five sounds in each pitch and speech-feedback interval condition for each participant. We performed a 3 × 5 repeated-measures ANOVA on rating scores with within-subject factors of pitch (high, neutral, and low) and speech-feedback interval (50, 200, 350, 500, and 650 ms). If the sphericity assumption for ANOVA was violated, we applied Greenhouse–Geisser correction.

References

1. Gallagher, S. Philosophical conceptions of the self: implications for cognitive science. *Trends. Cogn. Sci.* **4**, 14-21 (2000).
2. Haggard, P. Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196 (2017).
3. Blakemore, S. J., Oakley, D. A., & Frith, C. Delusions of alien control in the normal brain. *Neuropsychologia* **41**, 1058-1067 (2003).
4. Frith, C., Blakemore, S., & Wolpert, D. M. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Rev.* **31**, 357-363 (2000).
5. Frith, C., Blakemore, S. J., & Wolpert, D. M. Abnormalities in the awareness and control of action. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **355**, 1771-1788 (2000).
6. Gallagher, S., & Trigg, D. Agency and anxiety: Delusions of control and loss of control in schizophrenia and agoraphobia. *Front Hum Neurosci* **10**, (2016).
7. Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. Awareness of action in schizophrenia. *NeuroReport* **14**, 1081-1085 (2003).
8. Synofzik, M., Their, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain* **133**, 262-271 (2009).
9. Voss, M., Chambon, V., Wenke, D., Kühn, S., & Haggard, P. In and out of control: brain mechanisms linking fluency of action selection to self-agency in patients with schizophrenia. *Brain* **140**, 2226-2239 (2017).
10. Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain* **133**, 3104-3112 (2010).
11. Werner, J. D., Trapp, K., Wüstenberg, T., & Voss, M. Self-attribution bias during continuous action-effect monitoring in patients with schizophrenia. *Schizophr. Res.* **152**, 33-40 (2014).
12. Blakemore, S. J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol. Med.* **30**, 1131-1139 (2000).
13. Ford, J. M. Studying auditory verbal hallucinations using the RDoC framework. *Psychophysiology* **53**, 298-304 (2016).
14. Frith, C. D. The cognitive neuropsychology of schizophrenia. *Psychology press* (1992).
15. Blakemore, S. J., Wolpert, D., & Frith, C. Why can't you tickle yourself? *Neuroreport* **11**, R11-16 (2000).
16. Miall, R. C., & Wolpert, D. M. Forward models for physiological motor control. *Neural netw.* **9**, 1265-1279 (1996).
17. Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. Fine-tuning of auditory cortex during speech production. *Psychophysiology* **42**, 180-190 (2005).

18. Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology* **44**, 522-529 (2007).
19. Ford, J. M., & Mathalon, D. H. Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *Int. J. Psychophysiol.* **58**, 179-189 (2005).
20. Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Arch. Gen. Psychiatry* **64**, 286-296 (2007).
21. Whitford, T. J., et al. Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychol. Med.* **41**, 959-969 (2011).
22. Seal, M. L., Aleman, A., & McGuire, P. K. Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cogn. Neuropsychiatry* **9**, 43-72 (2004).
23. Jones, S. R., & Fernyhough, C. Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Conscious. Cogn.* **16**, 391-399 (2007).
24. Synofzik, M., Vosgerau, G., & Newen, A. Beyond the comparator model: a multifactorial two-step account of agency. *Conscious. Cogn.* **17**, 219-239 (2008).
25. Synofzik, M., Vosgerau, G., & Voss M. The experience of agency: an interplay between prediction and postdiction. *Front. Psychol.* **4**, (2013).
26. Johns, L. C., Gregg, L., Allen, P., & McGuire, P. K. Impaired verbal self-monitoring in psychosis: effects of state, trait and diagnosis. *Psychol. Med.* **36**, 465-474 (2006).
27. Johns, L. C., & McGuire, P. K. Verbal self-monitoring and auditory hallucinations in schizophrenia. *Lancet* **353**, 469-470 (1999).
28. Johns, L. C*, et al.* Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychol. Med.* **31**, 705-715 (2001).
29. Goldberg, T. E., Gold, J. M., Coppola, R., & Weinberger, D. R. Unnatural practices, unspeakable actions: a study of delayed auditory feedback in schizophrenia. *Am. J. Psychiatry* **154**, 858-860 (1997).
30. Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. Coercion changes the sense of agency in the human brain. *Curr. Biol.* **26**, 585-592 (2016).
31. Caspar, E. A., Lo Bue S., Magalhães De Saldanha da Gama P. A., Haggard, P., & Cleeremans, A. The effect of military training on the sense of agency and outcome processing. *Nat. Commun.* **11**, 4366 (2020).
32. Haggard, P., Clark, S., & Kalogeras, J. Voluntary action and conscious awareness. *Nat. Neurosci.* **5**, 382-385 (2002).
33. Moore, J. W., & Obhi, S. S. Intentional binding and the sense of agency: a review. *Conscious. Cogn.* **21**, 546-561 (2012).
34. Yoshie, M., & Haggard, P. Negative emotional outcomes attenuate sense of agency over voluntary actions. *Curr. Biol.* **23**, 2028-2032 (2013).

35. Yoshie, M., & Haggard, P. Effects of emotional valence on sense of agency require a predictive model. *Sci. Rep.* **7**, 8733 (2017).
36. Asai, T., & Tanno, Y. The relationship between the sense of self-agency and schizotypal personality traits. *J. Mot. Behav.* **39**, 162-168 (2007).
37. Farrer, C., et al. The angular gyrus computes action awareness representations. *Cereb. Cortex* **18**, 254-261 (2008).
38. Imaizumi, S., & Asai, T. My action lasts longer: Potential link between subjective time and agency during voluntary action. *Conscious. Cogn.* **51**, 243-257 (2017).
39. Imaizumi, S., & Tanno, Y. Intentional binding coincides with explicit sense of agency. *Conscious. Cogn.* **67**, 1-15 (2019).
40. Maeda, T., Kato, M., Muramatsu, T., Iwashita, S., Mimura, M., & Kashima, H. Aberrant sense of agency in patients with schizophrenia: forward and backward over-attribution of temporal causality during intentional action. *Psychiatry Res.* **198**, 1-6 (2012).
41. Maeda, T., et al. Reduced sense of agency in chronic schizophrenia with predominant negative symptoms. *Psychiatry Res.* **209**, 386-392 (2013).
42. Franck, N., et al. Defective recognition of one's own actions in patients with schizophrenia. *Am. J. Psychiatry* **158**, 454-459 (2001).
43. Suzuki, K., Lush, P., Seth, A. K., & Roseboom, W. Intentional binding without intentional action. *Psychol. Sci.*, 0956797619842191 (2019).
44. Legaspi, R., & Toyoizumi, T. A Bayesian psychophysics model of sense of agency. *Nat. Commun.* **10**, 4250 (2019).
45. Wolpe, N., Haggard, P., Siebner, H. R., & Rowe, J. B. Cue integration and the perception of action in intentional binding. *Exp. Brain Res.* **229**, 467-474 (2013).
46. Limerick, H., Moore, J. W., & Coyle, D. Empirical evidence for a diminished sense of agency in speech interfaces. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,* Association for Computing Machinery (2015).
47. Kluft, R. P. First-rank symptoms as a diagnostic clue to multiple personality disorder. *Am. J. Psychiatry* **144**, 293-298 (1987).
48. Ross, C. A., Miller, S. D., Reagor, P., Bjornson, L., Fraser, G. A., & Anderson, G. Schneiderian symptoms in multiple personality disorder and schizophrenia. *Compr. Psychiatry* **31**, 111-118 (1990).
49. Faul, F., Erdfelder, E., Lang A. G., & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175-191 (2007).
50. Asai, T., Sugimori, E., & Tanno, Y. Auditory verbal hallucination in schizophrenic patients and the general population: The sense of agency in speech. *Hallucinations: Types, Stages and Treatments*, 33-59 (2011).
51. Asai, T., Sugimori, E., & Tanno, Y. A psychometric approach to the relationship between

hand–foot preference and auditory hallucinations in the general population: Atypical cerebral lateralization may cause an abnormal sense of agency. *Psychiatry Res.* **189**, 220-227 (2011).

52. Asai, T., & Tanno Y. Why must we attribute our own action to ourselves? Auditory hallucination like-experiences as the results both from the explicit self-other attribution and implicit regulation in speech. *Psychiatry Res.* **207**, 179-188 (2013).

53. Sugimori, E., Asai, T., & Tanno, Y. Sense of agency over thought: External misattribution of thought in a memory task and proneness to auditory hallucination. *Conscious. Cogn.* **20**, 688-695 (2011).

Data availability

Data for each participant from Experiment 1, Control experiment for Experiment 1, and Experiment 2 are available in a spreadsheet file of Supplementary Data. For details, see README sheet of the file.

Acknowledgements

H.I. was supported by JSPS KAKENHI Grant Numbers 18H01098, 19H05725, and 19H01777. T.A. was supported by JSPS KAKENHI Grant Number 17K13971. H.I. and T.A. were also supported by "Research and development of technology for enhancing functional recovery of elderly and disabled people based on noninvasive brain imaging and robotic assistive devices," Commissioned Research of National Institute of Information and Communications Technology (NICT) and Japan Agency for Medical Research and Development (AMED) (grant JP18dm0307008). The authors are grateful to Dr Takaki Maeda (Keio University School of Medicine) for his insightful comments on this manuscript from the viewpoint of a psychiatrist. We would like to thank Hiroki Tarumi for his assistance with data collection and Ayuko Misu and Marina Sano for their helpful discussions on the experimental design.

Author contribution statement

R.O., T.A., and H.I. designed the study; R.O. collected the data in all of the experiments; T.A. and S.I. summarized the data of the previous studies, which were compared with the data in the current study. R.O. analyzed the data; R.O. and H.I. wrote the manuscript; T.A. and S.I. reviewed and approved the final version of the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.