Speaking with an alien voice:

Flexible sense of agency during vocal production

Matthias K. Franken[1], Robert J. Hartsuiker[1], Petter Johansson[2], Lars Hall[2], & Andreas Lind[2]

[1]Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

[2]Department of Philosophy, Lund University Cognitive Science, Lund University, Box 192, 221 00 Lund, Sweden

Corresponding author: Matthias K. Franken, matthias.franken@ugent.be

Word Count: 10062

Abstract

Speakers monitor auditory feedback during speech production in order to correct for speech errors. The comparator model proposes that this process is supported by comparing sensory feedback to internal predictions of the sensory consequences of articulation. Additionally, this comparison process is proposed to support the sense of agency over vocal output. The current study tests this hypothesis by asking whether mismatching auditory feedback leads to a decrease in the sense of agency as measured by speakers' responses to pitch-shifted feedback. Participants vocalized while auditory feedback was unexpectedly and briefly pitch-shifted. In addition, in one block, the entire vocalization's pitch was baseline-shifted ('alien voice'), while it was not in the other block ('normal voice'). Participants compensated for the pitch shifts even in the alien voice condition, suggesting that agency was flexible. This is problematic for the classic comparator model, where a mismatching feedback would lead to a loss of agency. Alternative models are discussed in light of these findings, including an adapted comparator model and the inferential account, which suggests that agency is inferred from the joint contribution of several multisensory sources of evidence. Together, these findings suggest that internal representations of one's own voice are more flexible than often assumed.

Keywords: sense of agency; speech production; pitch; voice; altered auditory feedback

Public Significance

This study suggests that the sense of agency during speech is more flexible than often assumed. The sense of agency is not simply the result of an absolute match between expected and observed sensory feedback. Instead, it may rely on the contribution of multisensory information and/or on temporal covariance between prediction and feedback.

Introduction

Auditory feedback plays a crucial role in speech production. Speakers monitor auditory feedback during speech production in order to correct for speech errors and avoid them in the future (Burnett et al., 1998; Guenther, 2016; Houde & Jordan, 1998; Houde & Nagarajan, 2011). This process is most commonly thought to be supported by a comparison between predicted and observed sensory feedback. In addition, some researchers have proposed that this comparison is also crucial in the generation of a sense of agency (Sarah-J. Blakemore et al., 1998; Korzyukov et al., 2017; Rauschecker, 2011), i.e. the sense that one is in control of one's speech output. The current paper further investigates the interplay between these two functions of comparing predicted and observed sensory input (i.e., error monitoring and sense of agency), and, more specifically, whether a large mismatch between predicted and observed sensory feedback leads to a decrease in the sense of agency, as suggested by the comparator model (Gallagher, 2000; Wolpert & Ghahramani, 2000).

It is well established that speakers compensate for unexpected perturbations of auditory feedback (Burnett et al., 1998; Franken et al., 2018; Guenther, 2016; Houde & Jordan, 1998), similar to corrections for feedback perturbations in other (non-speech) motor domains (Shadmehr & Mussa-Ivaldi, 1994). For example, in some of these studies, speakers received auditory feedback that was unexpectedly pitch-shifted in real time. This perturbation of auditory feedback tends to lead to compensatory responses. If the feedback's pitch was shifted up, speakers tended to compensate by decreasing the pitch in their production, while they increased their own pitch when the feedback's pitch was shifted down. Both in speech and in non-speech motor control, researchers have argued for a comparator model of error monitoring (Wolpert & Ghahramani, 2000). According to this framework, an internal forward model is generated during motor planning, which predicts the sensory consequences of the issued motor commands. This prediction is subsequently compared to the perceived sensory feedback. In case of a mismatch between predicted and observed feedback, an error signal can lead to compensatory responses and/or adaptation learning (Guenther, 2016; Houde & Nagarajan, 2011).

In addition, the comparator model of the sense of agency suggests that not only do forward models play a role in self-monitoring and error correction during speech production, but they also allow us to distinguish between self- and externally generated speech and thus in the generation of a sense of agency over one's vocal output (Sarah-Jayne Blakemore et al., 2002; Sarah-Jayne Blakemore & Frith, 2003; Frith, 2005; Wolpert & Ghahramani, 2000). The sense of agency is the experience of being the agent behind one's actions and their consequences (Haggard, 2017). According to Haggard (2017), a core aspect of the sense of agency is an association between a voluntary action and an outcome. In other words, it involves the recognition that a sensory event was under our control and associated with an intended action. If observed sensory input corresponds to predicted sensory feedback, it is labeled as self-generated and a sense of agency is generated. Thus, the comparator model suggests that internal forward models are crucial in the establishment of a sense of agency.

An alternative view on sense of agency is proposed by inferential models, which argue that agency is a more dynamic construct, arising from contextual inferences of the most likely cause of an action (Dennett, 1991; Wegner & Wheatley, 1999). While this inferential account does not deny that the comparison between predicted and observed sensory feedback plays a role in the sense of agency, it argues that this comparison is only one of multiple sources of information that jointly lead to a sense of agency. Crucially, this process may also include external cues, such as whether there are viable alternative sources of the sensory event (Wegner & Wheatley, 1999), or whether the sensory event corresponds to the subject's intention (Moore et al., 2009). This contrasts with the comparator model, which places the burden of establishing a sense of agency on internal motoric cues. Evidence in favor of the inferential account has been gathered from cases of abnormal authorship processing. A well-known example is the so-called rubber hand illusion (Botvinick & Cohen, 1998; Tsakiris et al., 2006), where participants experience an illusion of agency over a rubber hand after simultaneous and identical stimulation or movement of the rubber hand and their actual hand. This illusion suggests that the sense of agency is flexible and can, under some conditions, be associated with unintended movements. This is also in line with recent suggestions that the sense of agency is a complex, multi-layered and dynamic phenomenon (Moore, 2016; Synofzik et al., 2008), which depends on the integration of multiple cues, both internal and external (Moore et al., 2009).

Within the speech domain, most studies have taken a comparator model-based perspective and focused on the role of forward model sensory predictions in the vocal sense of agency (Korzyukov et al., 2017; Subramaniam et al., 2018). According to this framework, if the predicted feedback does not fully match the observed feedback, this can lead either to a response to compensate for the articulation error and/or adaptation of the forward model (where in both cases the auditory input is assumed to be self-caused), or to the rejection of the auditory input as self-produced (i.e., to a decreased or lost sense of agency). However, there have been few studies investigating how these two functions (error detection/correction and sense of agency) interact. Logically, in order to use auditory input for error-monitoring, the auditory input has to be recognized as feedback from one's own, self-produced speech and thus a minimal sense of vocal agency is required. From this perspective, it can be assumed that a loss or decrease of the sense of agency over auditory input will lead to fewer speech error corrections. In line with this assumption, in response to unexpected pitch shifts in auditory feedback, speakers tend to compensate less for large shifts (> 250 cents[1]) compared to smaller shifts (Scheerer et al., 2013). It has been argued that this is the result of the large shifts being unlikely to be self-generated, leading to less error correction (Hawco et al., 2009; Korzyukov et al., 2017; Scheerer et al., 2013). Therefore, these

[1] The unit 'cents' refers to a psychometric scale to express a frequency interval, where doubling the frequency in Hertz corresponds to a 1200 cents increase. An interval of 100 cents reflects the interval between, for example, C and C# in western music. See below (methods) for more information.

authors hypothesize that if the feedback perturbation is large enough, the perturbation leads to a loss of vocal agency. If it is small enough, a sense of vocal agency is still maintained and thus speakers may use the mismatching feedback for compensatory responses or adaptive learning. In the current study, we test whether a 500 cents pitch shift that is constant throughout production, leads to a loss of vocal agency as measured by the speakers' responses to unexpectedly pitch-shifts of different magnitudes. While the synthesis between comparator and inferential approaches proposed by Moore et al. (2009) suggests that these frameworks are not necessarily mutually exclusive, the results of the current study will allow us to further constrain current theoretical models of agency in speech.

From the perspective of inferential models of the sense of agency, several studies, for example using the rubber hand illusion, have suggested the sense of agency is flexible in limb movement (Caspar et al., 2015; Kalckert & Ehrsson, 2012; Tsakiris & Haggard, 2005). However, it is difficult to investigate whether the sense of agency is also flexible in speech production. Lind et al. (2014) investigated this using a Stroop task. Participants performed a Stroop task, while in some trials, their auditory feedback was replaced with a recording of their own voice producing the incorrect response. Participants accepted recordings of erroneous answers as their own, self-produced feedback, although in reality they produced the correct response. In addition, Zheng et al. (2011) replaced participants' auditory feedback with recordings of a stranger's voice producing the same word. In an explicit judgment task, the participants sometimes accepted the stranger's voice as their own voice, indicative of a sense of agency. Similar results were obtained when participants were immersed in a virtual reality environment (Banakou & Slater, 2014, 2017; Tajadura-Jiménez et al., 2017). In these studies, participants also aligned the pitch in their own speech production with the voice they had heard. It has been argued that this speech alignment may be indicative of a sense of agency, given that synchronous visuomotor feedback in a virtual reality environment led to increased agency (as measured by explicit survey questions) and pitch alignment, while asynchronous visuomotor feedback did not (Banakou & Slater, 2014).

The sense of agency has proven to be difficult to measure. While, as we have seen, there have been a few fruitful attempts to investigate sense of agency through implicit measures, most studies have made use of questionnaires. Here, participants are explicitly asked to rate their sense of agency, to what extent they felt in control, or whether they thought the sensory event was caused by themselves. Although insightful, the explicit nature of these survey questions makes their results hard to interpret, as there is often no clear baseline measurement (Lush, 2020). In addition, it has been argued that participants' explicit judgement of agency should be distinguished from an implicit feeling of agency (Synofzik et al., 2008). Saito et al. (2015) found no correlation between explicit and implicit measures of agency, suggesting that the feeling of agency and the explicit judgement of agency are distinct.

In the current study, we present a novel implicit way of testing the flexibility of the sense of agency in a speech production task. Participants will be exposed to unexpected pitch changes in auditory feedback during speech production. Specifically, while participants vocalize for three seconds, auditory feedback will be intermittently pitch-shifted at jittered onsets relative to speech onset. Based on the previous research discussed above, these short, unexpected pitch shifts are expected to lead to compensatory responses in the following way: when feedback pitch is unexpectedly shifted up, participants respond by decreasing their pitch, and vice versa. These pitch shifts will vary in direction (sudden pitch increases or decreases) and magnitude (100 cents or 400 cents). Also based on previous research, we expect the 100 cent shifts to generate larger compensations compared to the 400 cent shifts. Aside from these short, unexpected pitch shifts, the feedback voice will either sound like their own voice (normal voice) or will be manipulated by a constant large pitch increase of 500 cents (alien voice). The latter condition thus makes the feedback sound unlike their own voice. From a comparator model's perspective, this alien voice feedback should lead to a loss of sense of agency, as feedback pitch manipulations of over 250 cents have previously been shown to reduce agency, due to the mismatch with internal sensory predictions. The current study will investigate to what extent this large and constant feedback perturbation (i.e., the alien voice) affects the responses to the brief unexpected pitch changes. In other words, we will consider compensatory responses to brief unexpected pitch shifts in auditory feedback as an implicit measure of the sense of agency over vocal output. This interpretation is in line with previous findings suggesting an association between response magnitude and the sense of agency (Subramaniam et al., 2018). Furthermore, Koryukov et al. (2017) found that ERP responses to 100 cents pitch shifts are qualitatively different from ERP responses to 400 cents. Specifically, they found that only cortical responses to the 400 cents shifts were sensitive to pitch shift direction. The authors argued that this qualitative difference could be taken as an index of the sense of agency. Finally, others have reported that predictable pitch shifts lead to smaller responses compared to unpredictable ones (Behroozmand et al., 2012; Scheerer & Jones, 2014), which is also in line with an association between response magnitude and the sense of agency. If the alien voice feedback reduces the sense of vocal agency – as predicted by the comparator model's emphasis on internal forward models - smaller compensatory responses are expected, compared to normal voice feedback. On the other hand, from an inferential model's perspective, if agency mainly relies on external cues, the alien voice should only have minimal impact on the sense of agency. Although the feedback is pitch-shifted, it is still the only auditory input that matches the speaker's articulation in all respects (timing, spectral characteristics), except for pitch. Therefore, the inferential model predicts that even with alien voice feedback, contextual cues will still engender a sense of agency, and thus similar compensatory responses should occur with both normal voice and alien voice feedback. In addition, if pitch alignment is indicative of agency as suggested in some studies discussed above (Banakou & Slater, 2017; Zheng et al., 2011), this account predicts that speakers will align their pitch with the alien voice. These hypotheses will be tested in two experiments, one where the feedback voice manipulation is blocked (Experiment 1), and one in which it was randomized (Experiment 2).

Experiment 1

Methods

Participants

Thirty-six volunteers (10 male and 26 female participants; average age = 22.7, SD = 3.13) participated in the experiment in exchange for either course credit or payment. All participants were native speakers of Dutch and had no history of speech or hearing impairment. The experiment was approved by the local ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University. All participants signed a consent form in accordance with the declaration of Helsinki.

Previous studies showed large effect sizes for detecting that the response magnitude to 100 cents is larger compared to 400 cents pitch shifts (Korzyukov et al., 2017; Subramaniam et al., 2018). In order to account for possible small sample biases, a power analysis was carried out to detect a somewhat smaller effect size in a one-tailed dependent t test (Cohen's $d = 0.7$). In order to achieve 95% power, a sample size of 24 participants would be needed. We decided to take a larger sample (N=36) as we did not have a clear estimate of the effect size of interest. With this sample size, we had 95% power to detect an effect size of Cohen's $d = 0.56$.

Experimental Procedure

Each trial, the appearance of the letters "EE" (pronounced in Dutch as /e/) on a computer screen provided a signal for participants to start vocalizing the vowel /e/ and to hold the vowel until the letters disappeared after 4s. Participants were instructed to try to keep the volume, pitch, and articulation of the vowel constant. During vocalization, participants received real-time auditory feedback of their own voice through the headphones (see Figure 1 for an illustration of the experimental design). The experiment consisted of a normal voice block and an alien voice block. In the normal voice block, pitch was shifted during each vocalization either upwards, downwards, or not at all. This happened three times during each vocalization, and each pitch shift lasted 200ms. No other acoustic parameter was manipulated. In the first half of each experimental block, the absolute magnitude of the shifts was either 100 cents or 400 cents, and the other magnitude was used in the second half of the blocks. The order of shift magnitudes was counterbalanced across participants. The shifts were separated from each other and from speech onset by a jittered interval of 600 to 800ms. The pitch shifts were randomized throughout the block of trials in such a way that each set of two consecutive trials contained two upward pitch shifts, two downward pitch shifts, and two null shifts. A block of trials would consist of 100 trials and thus of 300 shifts, including 50 shifts of each perturbation type as well as 100 null shifts. In the alien voice block, the same brief pitch shifts were applied, on top of a constant +500 cents pitch shift that lasted for the entire block. A demo of how the alien voice feedback would sound is available in the supplementary materials.

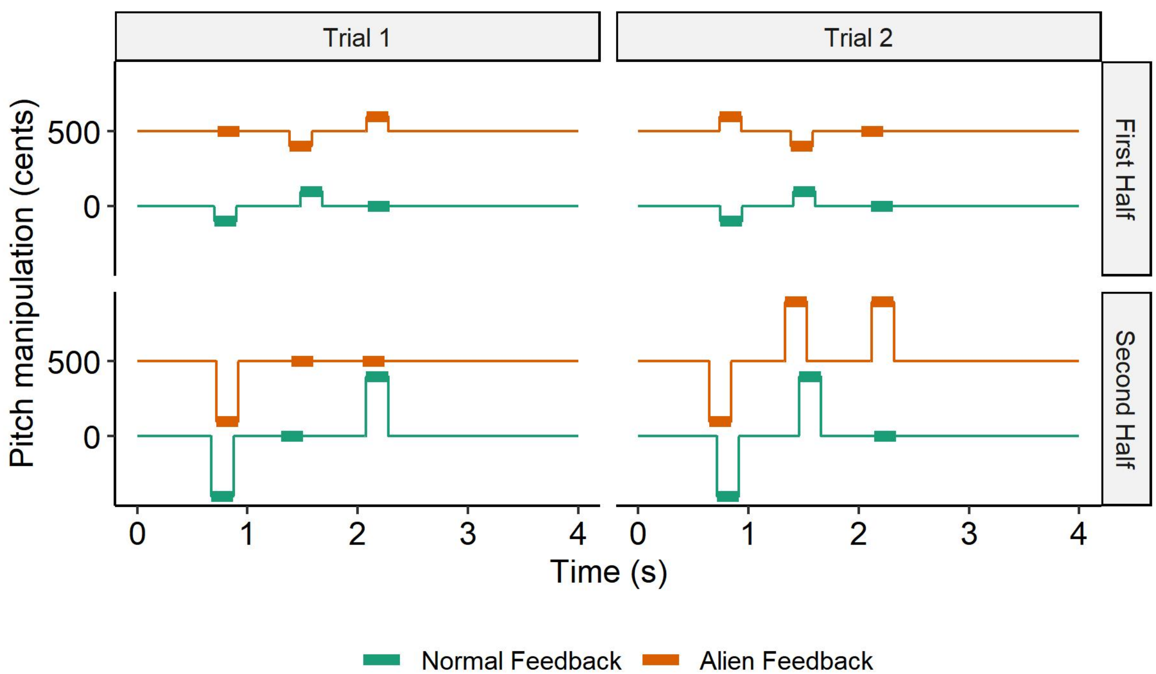

Figure 1. Experimental design. Example of two consecutive trials, for both halves of the alien voice block and the normal voice block. The y axis indicates the feedback pitch manipulation,

which were -100, +100 or 0 cents in one half of the block, and -400, +400 or 0 cents in the other half of the block. Every set of two consecutive trials contained two pitch shifts of each of these three values.

Equipment

Participants wore a custom-built pair of headphones built by the last author (Franken et al., 2019) with a microphone boom to which a DPA microphone (DPA 4088-B) was attached, positioned at about 2cm from the participant's mouth. The microphone was connected to a Xenyx 802 audio mixer, which sent the signal to an Eventide Eclipse multi-effects processor, which generated the pitch manipulations. The delay between the time that an acoustic signal is picked up by the microphone, and the corresponding signal output from the headphones, was about 10ms. The pitch manipulations were controlled via MIDI by a custom PureData (Puckette, 1996) program written by the first author. The output signal from the multi-effects processor was sent, via a different channel on the Xenyx 802 audio mixer to an Aphex HeadPod 4 headphones amplifier to the headphones. At the same time, both the microphone signal and the manipulated audio signal were sent to a MicroBook IIc audio interface connected to the laptop in order to store them for offline analysis. The volume of the auditory feedback was set 10dB above the signal picked up by the microphone (as measured with the Bruel & Kjaer 2250 sound level meter), as in some previous studies, in order to drown out bone-conducted feedback (Behroozmand et al., 2014; Hawco et al., 2009; Liu et al., 2011).

Rating Questionnaire

In order to assess the conscious experience of the alien voice feedback, participants filled out a questionnaire after the entire experiment. The questionnaire consisted of 9 statements to be rated on a 6-point scale (ranging from "helemaal niet akkoord", *completely disagree*, to "helemaal akkoord", *completely agree*). The statements are found in **Error! Reference source not found.** and were based on similar questionnaires used in studies on the rubber hand illusion (Botvinick & Cohen, 1998; Kalckert & Ehrsson, 2012). All statements were rated once for each block (the alien voice and the normal voice block), but both ratings were done after both blocks in order to not alert the participant in between blocks that we were interested in the sense of agency. A limitation of this approach is that, depending on the block order, the rating for one block had to rely on memory of events somewhat further in the past, while the rating for the other block was done right after the block ended.

Table 1. Questionnaire statements for the rating task, their translation, and their abbreviation to be used in the remainder of the article.

Questionnaire statements	Translation	Abbreviation
Ik hoorde mijn eigen stem in de koptelefoon.	I heard my own voice in the headphones.	Own_voice
De stem die ik hoorde klonk als een gemanipuleerde versie van mijn stem.	The voice I heard sounded like a manipulated version of my voice.	Manip_voice
De stem die ik hoorde was van iets/iemand anders.	The voice I heard belonged to something/someone else.	Other_voice
Ik hoorde twee verschillende stemmen tegelijkertijd.	I heard two voice at the same time.	Two_voices
Ik voelde mezelf in controle over het geluid in de koptelefoon.	I felt in control over the sound in the headphones.	Control

Na verloop van tijd leek de stem meer op mijn stem dan aan het begin.	After a while, the voice was more similar to my voice than at the beginning.	Time_voice
Het voelde alsof ik het geluid in de koptelefoon produceerde.	It felt like I produced the sound in the headphones.	Self_produced
Mijn articulatie was stabiel.	My articulation was stable.	Artic_stable
Als ik iets veranderde in mijn articulatie, veranderde de klank in de koptelefoon op dezelfde manier.	If I changed my articulation, the sound in the headphones changed in the same way.	Artic_change

For the analysis of the questionnaire ratings, an exploratory factor analysis was performed to examine the data structure. The number of factors to extract was determined based on a parallel analysis implemented in the R package 'psych'. Factors were deemed relevant when their Eigen value was larger than the corresponding Eigen value of factors based on simulated data or on resampling of the data (see Results). The factor analysis was carried out using a maximal likelihood approach with oblimin rotation.

Analysis

All data and analysis scripts are publicly available via doi.org/10.17605/OSF.IO/86JD5. Sometimes vocalization was too soft or initially too soft to trigger the pitch shifts in time. This sometimes led to mistiming of the pitch shifts. As long as the pitch shifts were applied during vocalization with ample time of vocalization around them (200ms before and 700ms after shift onset), the data were included in the analysis. A pitch estimation algorithm based on autocorrelation in Praat

(Boersma & Weenink, 2017) was used to estimate pitch in Hertz in every vocalization with a 1ms resolution. The resulting pitch contours were exported to Matlab (R2016b). From every perturbation's pitch contour (including null perturbations), epochs were extracted from 200ms preceding to 700ms following the perturbation onset. Pitch was converted to the cents scale as follows:

$$pitch_{cents} = 1200 \cdot \log_2\left(\frac{pitch_{Hz}}{baseline_{Hz}}\right)$$

Here, $\text{baseline}_{\text{Hz}}$ is the mean pitch in Hertz over the 100ms preceding perturbation onset. The pitch contours for all epochs were visually inspected for pitch estimation errors. As a result of visual inspection, epochs with sharp discontinuities or unusually high variability were discarded. Epochs where more than 10% of the pitch contour was undefined (due to a pitch estimation failure) were discarded as well. On average, this procedure led to a remaining 536.7 trials per participant (89.4%). Undefined stretches in remaining epochs' pitch contour were linearly interpolated from neighboring samples. For every participant, block, and perturbation condition, the average pitch response contour was calculated by averaging across trials. Only conditions with at least 20 trials for a participant were included in further analysis. This was the case for all conditions in all participants. In order to make sure response magnitude estimations were not affected by gradual drifts, difference waves were calculated by subtracting the average for the null perturbation from the average of the corresponding non-null perturbations (of the same participant). The sign of the difference waves for the upward perturbations was flipped, so positive values indicate opposing responses while negative values indicate following responses. For every participant, voice condition, and perturbation, the compensation response magnitude was estimated as the maximal value between 60ms after perturbation onset and the end of the epoch (700 ms after perturbation onset). This window was chosen as 60ms is considered to be the minimal time necessary to respond to an unexpected pitch shift (Chen et al., 2007; Larson et al., 2001).

In addition, we used a trial classification method to classify every single trial as containing either an opposing or a following response. The trials were classified based on the slope of the pitch contour over a 100ms-300ms time window after perturbation onset (Franken et al., 2018). This analysis is reported in the supplementary materials.

Finally, an additional analysis looked at how speakers changed pitch over the course of an experimental block, as a function of the feedback voice. For every vocalization, the average pitch value from 50ms to 100ms after vocalization onset was extracted from the pitch contours. Subsequently, these pitch values were averaged across participants for each trial, as further detailed in the results section.

Statistical inference

In order to assess whether pitch shifts led to a compensatory response, each condition's response contour was compared to the response contour for the 0 cents shift with the same voice. This comparison was carried out using a cluster-based permutation test (Maris & Oostenveld, 2007). For every condition, a t-value was calculated at each time sample, and neighboring time points that exceeded a value corresponding to an uncorrected p value of 0.05 were clustered. The summed t value was calculated per cluster and the largest sum was used as the statistic of interest. The same was done after permuting condition labels randomly, arriving at a permutation distribution against which the original statistic value was tested. Cohen's d values were calculated across the time window from the start to the end of the largest cluster.

Whereas the cluster-based permutation tests investigate for each condition separately whether a compensatory response was generated in response to the pitch shift, linear mixed effects modeling was used for investigating the effects of pitch shift magnitude, pitch shift direction, and feedback voice condition on the response magnitude. All these statistical tests were carried out in R (R Core Team, 2018). Response magnitudes were modeled as a function of voice, perturbation magnitude, and perturbation direction in linear mixed effects models. We started out with the maximal model (including random slopes for all main and interaction terms). If this model did not converge, first the correlations between random effect factors were removed. If the model still did not converge, the most complex random interaction terms were removed one by one, until the model converged. Reported p values are calculated using Satterthwaite's methods for estimating degrees of freedom. The omnibus results shown are a type-III table of variance calculated using the anova() function in R, while the pairwise comparisons are calculated using the 'emmeans' package, with Tukey-adjusted p-values if appropriate. All mixed effects modelling was performed using the R packages 'lme4' (Bates et al., 2015), 'afex' (Singmann et al., 2019) and 'lmerTest' (Kuznetsova et al., 2016).

Results

Figure 2 shows the compensation responses to the unexpected pitch shifts. A cluster-based permutation test was used to compare every condition to the corresponding 0 cents shift condition. In all but one condition, the response contour significantly differed from the response to a 0 cents shift (Table 2). Only in the normal voice -400 cents shift, the difference was not significant, although a trend in the same direction can be seen in Figure 2. In addition, the data in Table 2 suggests that the responses to 400 cents shifts started later compared to responses to 100 cents shifts. Although not explicitly tested, similar trends can be found in figures shown in earlier studies (Korzyukov et al., 2017; Subramaniam et al., 2018).

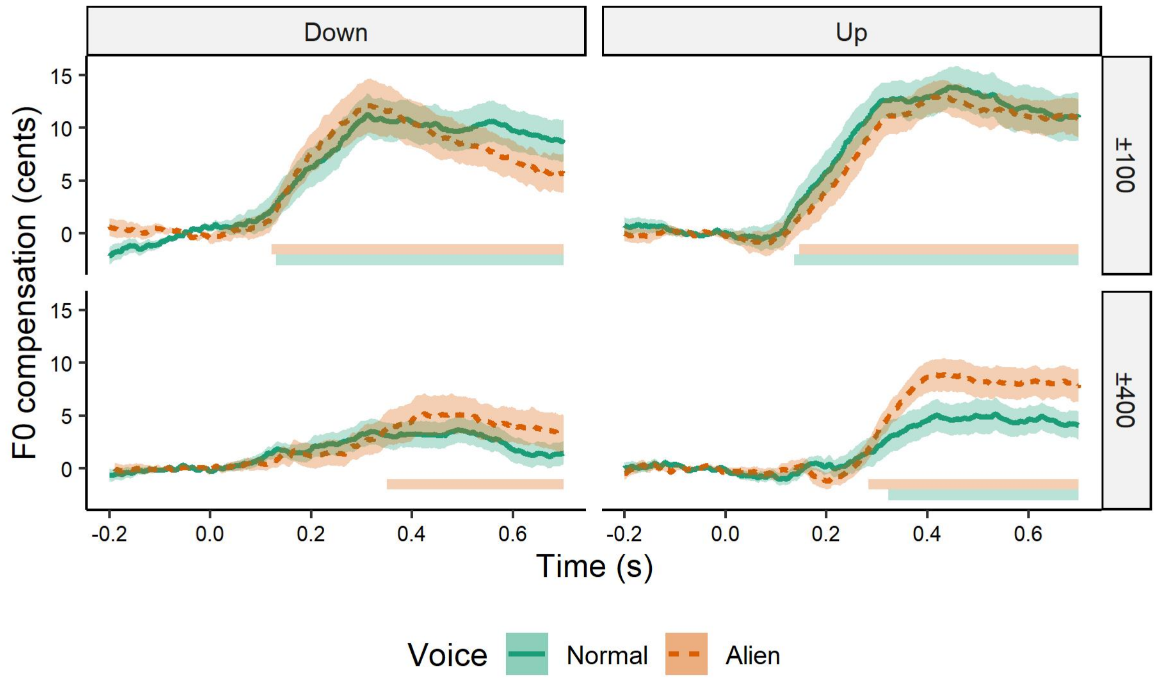

Figure 2. Average pitch compensation contours as a function of feedback voice, perturbation direction, and perturbation magnitude. Horizontal bars at the bottom of each panel indicate the extent of the largest cluster driving the difference between this condition and the corresponding zero shift perturbation.

Table 2. Results of cluster-based permutation tests

Voice	Pitch Shift (cents)	Onset largest cluster (ms)	p	Cohen's d
Normal	100	136	< .001*	1.27
Normal	-100	131	< .001*	-0.96
Normal	400	323	.0076*	0.50
Normal	-400		.150	-0.37
Alien	100	146	< .001*	1.53

Alien	-100	122	< .001*	-1.06
Alien	400	283	< .001*	0.93
Alien	-400	350	.010*	-0.52

Next, we examined the magnitude of the response's peak, which was modeled as a function of voice type, shift direction, and shift magnitude in a linear mixed effects model including random slopes for all main effects and two-way interactions, but not the three-way interaction between voice, perturbation magnitude, and perturbation direction (Figure 3). The fixed effects are shown in Table 3. There was a main effect of shift magnitude, with larger responses for ±100cents shifts than for ±400 cents shifts (est. = 5.83, $t(35) = 5.89$, $p < .001$). None of the other terms were significant, suggesting that whether the feedback was the participant's normal voice or the alien voice did not affect the magnitude of participant's responses.

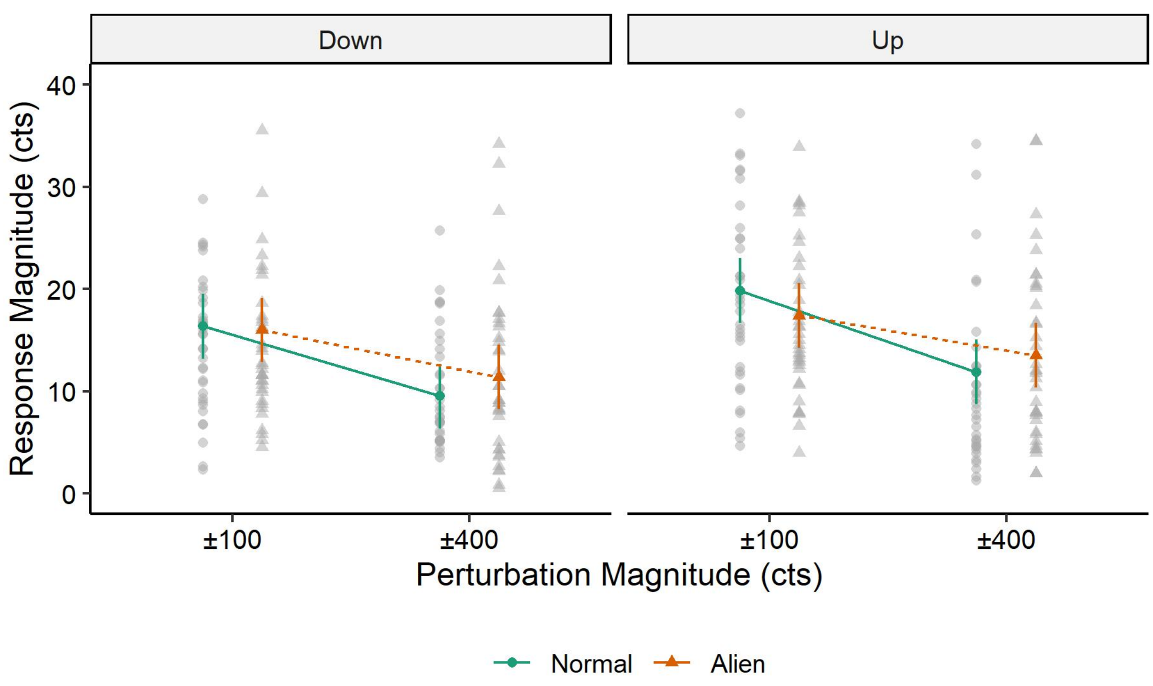

Figure 3. Response magnitude as a function of perturbation magnitude, feedback voice, and perturbation direction.

Table 3. Fixed effects of the model Response Magnitude ~ Voice*ShiftMagnitude*ShiftDirection + (1 + Voice + ShiftMagnitude + ShiftDirection + Voice:ShiftMagnitude + Voice:ShiftDirection + ShiftMagnitude:ShiftDirection | Participants)

	SS	df	F	p
Voice	1.66	1, 35	0.030	.86
ShiftMagnitude	**1929.72**	**1, 35**	**34.666**	**< .001***
ShiftDirection	193.09	1, 35	3.469	.071
Voice:ShiftMagnitude	173.13	1, 35	3.110	.087
Voice:ShiftDirection	17.22	1, 35	0.309	.582
ShiftMagnitude:ShiftDirection	0.35	1, 35	.006	.937
Voice:ShiftMagnitude:ShiftDirection	14.73	1, 35	0.265	.610

The response to large shifts was reduced compared to small pitch shifts (by on average 5.83 cents, SE = 0.99 cents), but we do not find an effect of voice type. This suggests that voice type did not affect vocal agency, in contrast with predictions from the comparator model's perspective. In order to investigate this more closely, we examined participants' responses on the questionnaires. Panels A and B in Figure 4 show how participants rated the different statements (see Table 1) for each block. Every statement was rated by every participant for the normal voice block and for the alien voice block separately. Panel A shows boxplots for the raw rating values for each statement in each voice condition. Panel B shows, for each statement, the difference in rating between blocks, where negative values indicate a lower rating for the Alien Voice compared to the Normal Voice. In order to reduce data dimensionality, we carried out an exploratory factor analysis. The scree plot in panel C of Figure 4 suggests that four factors should be sufficient to capture the data. Table 4 shows how these four factors load on the nine statements. Based on these results, factors 1 and 4 seem most relevant for our purposes. Factor 1 loads mostly on the statements own_voice, manip_voice, and other_voice, suggesting that a high score on Factor 1 is associated with a higher agreement that the voice is one's own voice, and a lower agreement that the voice is a manipulated version of one's own voice or another person's voice. Thus, this factor seems to reflect how similar the feedback voice was perceived to be to one's own voice. Factor 4 loads mostly on the statements control, time_voice, self_produced, and artic_change, suggesting that a high score indicates the speaker feels they were in control of the feedback, that the feedback sounded more like their own voice over time, was the result of their own production, and corresponded to their articulation. Thus, factor 4 seems to reflect how well the feedback corresponded to their own speech production. On the other hand, factors 2 and 3 loaded most heavily on a single statement each, which seem less relevant for the sense of agency (two_voices, artic_stable). This is confirmed by the results in panel D of Figure 4, where boxplots show each factor's score as a function of voice condition. Pairwise t-tests suggest that factors 1 and 4 score higher in the normal voice compared to the alien voice (Factor 1: $t(35) = 4.48$, $p < .001$, *Cohen's d* = 1.05; Factor 4: $t(35) = 2.80$, $p = .025$, *Cohen's d* = 0.48). There were no differences between voice conditions for the other two factors (Factor 2: $t(35) = 0.20$, $p > .99$, *Cohen's d* = 0.03; Factor 3: $t(35) = 0.33$, $p > .99$, *Cohen's d* = 0.05; all p-values Holm-corrected for a family of four tests). These results suggest that the alien voice reduced the sense of agency as measured by explicit judgments.

The results shown in Figure 4 suggest that the statement 'the voice in the feedback sounds like my own voice' was the most affected by voice condition. Ratings on this statement also showed the largest interindividual variability with respect to the effect of feedback voice, and factor 1 was very strongly loaded on especially this statement. Table 5 shows the Spearman rank correlation coefficients for all pairwise comparisons between statements. These show that participants who agreed with the feedback being their own voice less in the alien voice than in the normal voice, also thought the alien voice was more likely to be a manipulated version of their voice or even another person's voice. They also felt less like the alien voice was self-produced, and tended to feel less in control over the alien voice compared to the normal voice.

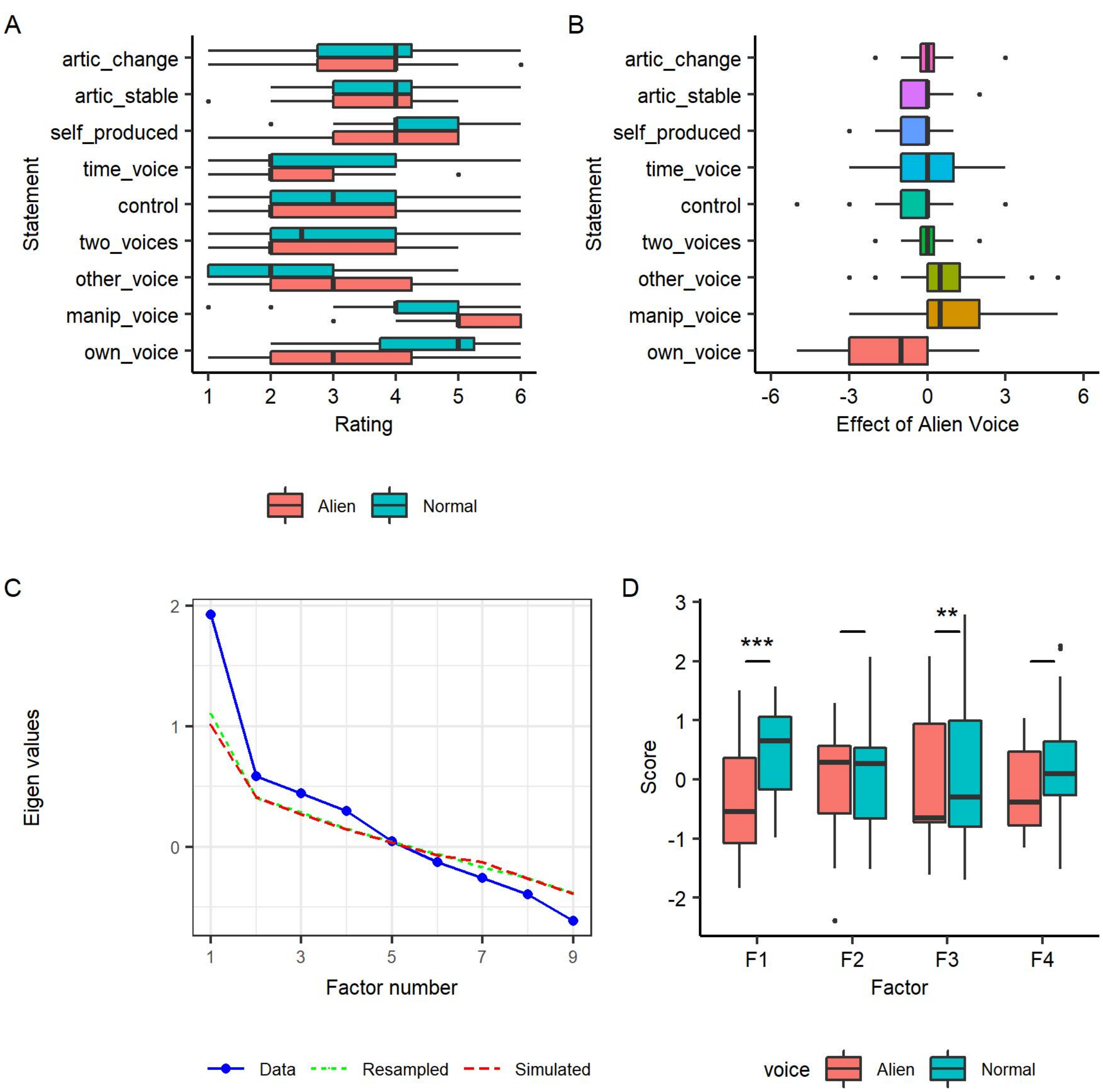

Figure 4. A. Boxplots in panel A show the ratings for each statement as a function of feedback voice. B. Boxplots in panel B show, for each statement, the difference in rating between the alien voice block and the normal voice block. Positive values indicate a higher rating for the alien voice block. C. A scree plot based on a parallel analysis in order to identify the number of factors that should be extracted for the exploratory factor analysis. D. Scores for each of the four factors from the exploratory factor analysis (see Table 4), as a function of feedback voice. The abbreviations for the different statements are used as in Table 1.

Table 4. Factor loadings for the four factors extracted by the exploratory factor analysis. Factor analysis was carried out by maximal likelihood estimation, with *oblimin* rotation. Only factor loadings above 0.1 are shown. The factors are ordered according to the total variance explained by each factor.

Statement	Factor 1	Factor 3	Factor 4	Factor 2
Own_voice	0.986	0.105		
Manip_voice	-0.452		-0.177	0.162
Other_voice	-0.611	0.169		-0.189
Two_voices		0.988		
Control	0.128		0.661	
Time_voice	-0.153	0.390	0.509	0.191
Self_produced	0.262	-0.213	0.402	-0.163
Artic_stable				0.993
Artic_change			0.483	

Table 5. Spearman rank correlation coefficients for the effect of alien voice feedback on statement ratings. Bold font and * indicate significance at the Holm-corrected 0.05 alpha level, while (*) indicates significance at the uncorrected 0.05 alpha level.

	Own_voice	Manip_voice	Other_voice	Two_voices	Control	Time_voice	Self_produced	Artic_stable
Manip_voice	**-.61***							
Other_voice	**-.77***	**.58***						
Two_voices	-.11	.14	.03					
Control	.49(*)	-.32	-**.52***	-.02				
Time_voice	-.01	-.19	-.04	.28	.33			
Self_produced	**.57***	-.28	-.50(*)	-.09	.39(*)	.18		
Artic_stable	.04	.03	-.14	-.12	.36(*)	.18	.08	
Artic_change	-.09	.07	.00	.03	.17	.21	.13	.18

As the ratings were only carried out at the very end of the experiment, participants would have to rely on memory to rate their experience of the first block, but less so for the second block. Closer analysis of the participants that experienced the alien voice block after the normal voice suggested that the pattern of results as shown in *Figure 4*, panel B is stronger in that group compared to participants who experienced the alien voice block first (Figure S4).

The rating data suggest that most of the variability that is relevant for the feedback voice difference can be captured by the 'own_voice' statement. As the results of this statement showed most variability, and factor 1 loaded heavily on this statement (factor loading = 0.986, Table 4), the participants were split in two groups based on their rating for this statement. The first group indicated, as expected, that the alien voice sounded less like their own voice than the normal voice (hereafter [OwnVoice-]), while the second group did not indicate this (hereafter [OwnVoice+], see Figure 5). Group assignment based on the 'own_voice' rating was independent of block order at the .05 alpha level ($\chi^2(1) = 1.83$, $p = .18$).

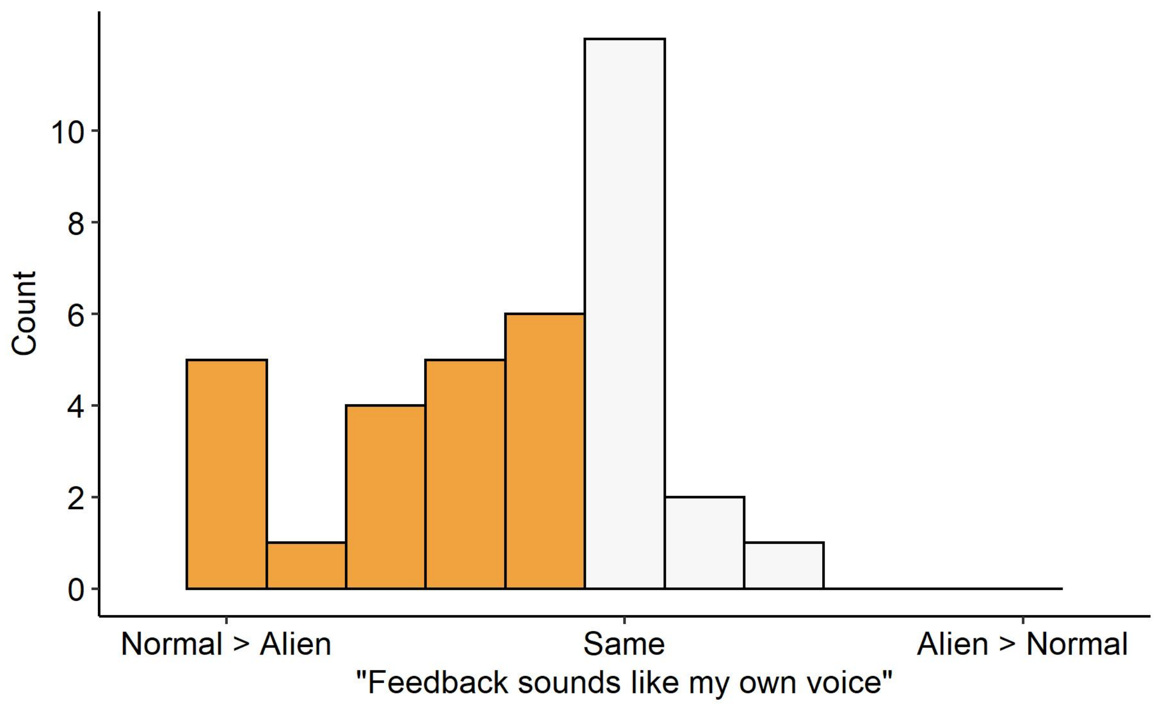

Figure 5. Histogram of participants' ratings on the first statement (own_voice). Coloring indicates the group split based on these results ([OwnVoice-] in orange, [OwnVoice+] in white).

Due to this individual variability in how the alien voice condition was perceived, it could be the case that the overall results presented earlier are obscuring group-specific effects. Therefore, we ran another linear mixed effects model, modeling response magnitude as a function of group (coloring in Figure 5), perturbation size, perturbation direction, and voice type. The results are in Table 6.

Table 6. Fixed effects results of the model Response Magnitude ~ Voice * ShiftMagnitude * Group * ShiftDirection + (Voice* ShiftMagnitude + ShiftDirection + ShiftDirection:Voice + ShiftDirection:ShiftMagnitude || Participants)

	SS	df	F	p
Voice	15.72	1, 34	0.275	.604
ShiftMagnitude	**2159.75**	**1, 34**	**37.739**	**< .001***
Group	69.81	1, 34	1.22	.277
ShiftDirection	225.13	1, 34	3.934	.0555
Voice:ShiftMagnitude	140.88	1, 68	2.462	.121
Voice:Group	**246.94**	**1, 34**	**4.315**	**.0454***
ShiftMagnitude:Group	122.55	1,34	2.141	.153
Voice:ShiftDirection	10.25	1, 34	0.179	.675
ShiftMagnitude:ShiftDirection	0.01	1, 34	0.0001	.991
Group:ShiftDirection	47.31	1, 34	0.827	.370
Voice:ShiftMagnitude:Group	60	1, 68	1.048	.310
Voice:ShiftMag:ShiftDir	15.89	1, 68	0.278	.600
Voice:Group:ShiftDirection	30.4	1, 34	0.531	.471
ShiftMag:Group:ShiftDir	16.5	1, 34	0.288	.595
Voice:ShiftMag:Group:ShiftDir	1.46	1, 68	0.026	.873

There was a clear main effect of shift magnitudeu, as in the initial overall analysis, as well as a barely significant interaction between Voice and Group, showing that the effect of Voice differed between groups (Figure 6). None of the other interactions were significant. The interaction seems to be driven by a significant difference in response magnitude to pitch shifts in the alien voice between groups, where participants who showed a difference between voices in their rating results ([OwnVoice-]) had smaller responses to pitch shifts in the alien voice compared to [OwnVoice+] participants, who did not show the rating difference (est. = -3.865, $t(55.3) = -2.011$, $p = .049$). Such a difference was not present in the normal feedback voice condition (est. = 0.214, $t(55.3) = 0.11$, $p = .912$). Although this effect was only barely significant, and based on a median split of a measure we do not know the reliability of, it suggests that the alien voice may have led to smaller responses in the [OwnVoice-] group but not in in the [OwnVoice+] group.

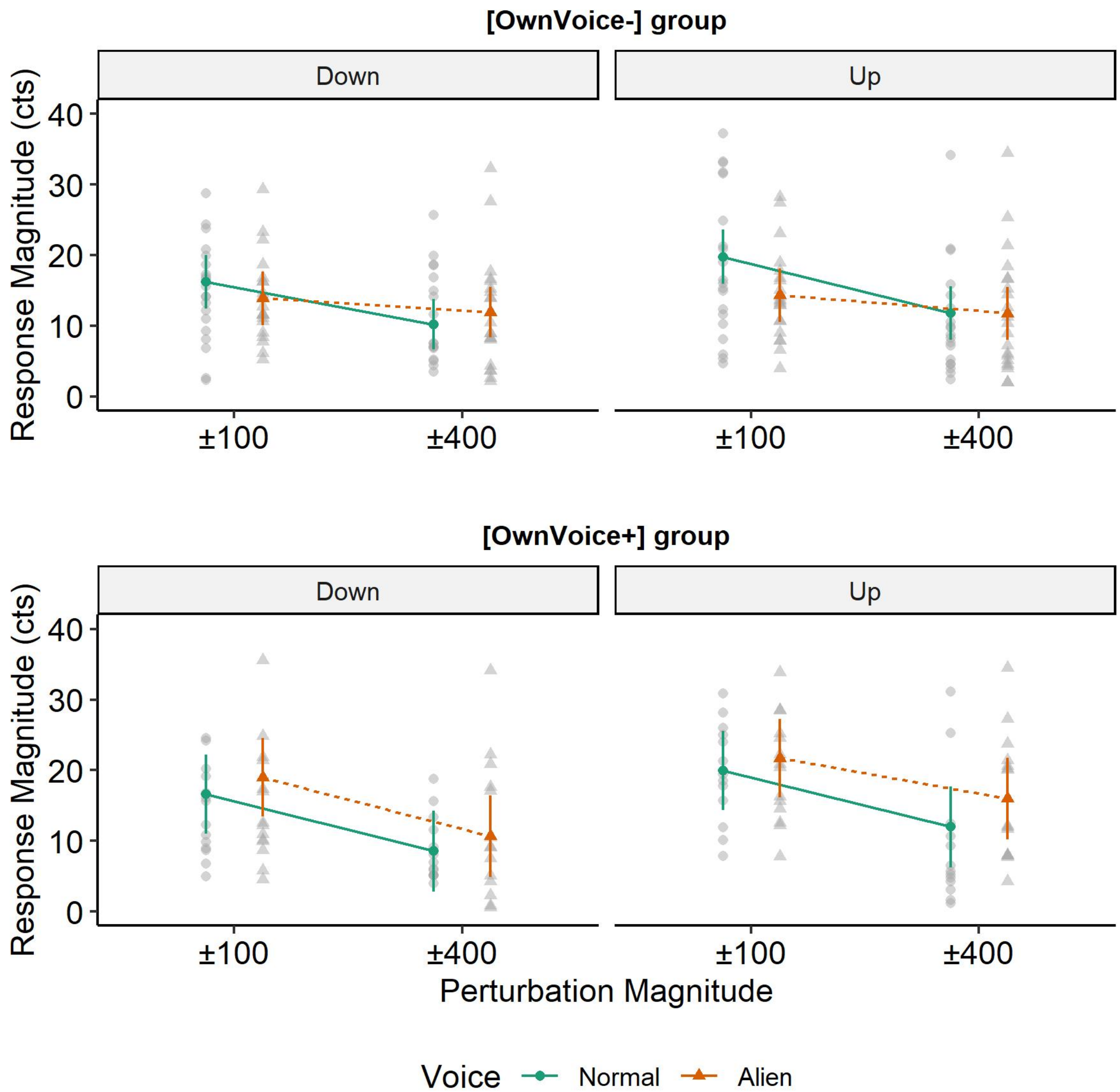

Figure 6. Response Magnitude as a function of pitch shift magnitude, pitch shift direction, feedback voice, and participant group.

In addition to participants' responses to the short pitch shifts, an analysis was performed to investigate how exposure to the alien voice feedback (i.e., a constant +500 cents pitch shift) affected participants' pitch production over time. A linear mixed effects model analysis suggested that the mean pitch from 50ms to 100ms after onset of a vocalization was significantly affected by the interaction between trial number and feedback voice ($F(1, 7087) = 12.45$, $p < .001$), suggesting that over the course of an experimental block, participants increased pitch more in the alien voice block compared to the normal voice block (by on average 0.21 cents per trial, or 21 cents over the course of 100 trials; Figure 7). In other words, participants tended to align their pitch with the alien voice feedback. It is unclear why participants tended to increase pitch over the course of a normal voice block as well. Interestingly, when a similar analysis was performed separately for each participant group, the results suggest that mainly the [OwnVoice-] participants show pitch alignment, whereas [OwnVoice+] participants did not show evidence of pitch alignment with the feedback (Figure 8). Although the three-way interaction between participant group, trial number, and feedback voice was not significant ($F(1, 7085) = 1.19$, $p = .28$), separate analyses showed an effect only for the [OwnVoice-] group. Compared to the normal voice, [OwnVoice-] participants increased pitch in the alien voice more strongly by 0.26 cents per trial. The difference in pitch increase between voice conditions for the [OwnVoice+] participants was only 0.13 cents per trial.

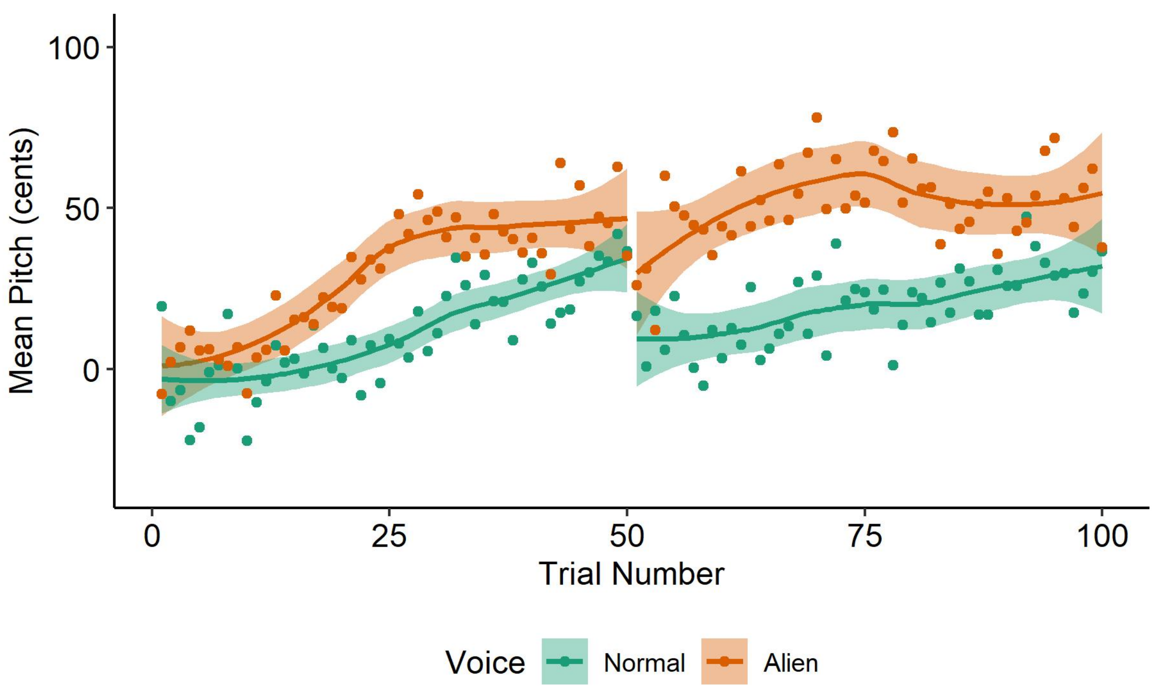

Figure 7. Change in mean pitch over the course of an experimental block as a function of feedback voice condition. The change in pitch is expressed as the change from the average pitch in the first 10 trials of a block. Participants got a break after 50 trials in each block.

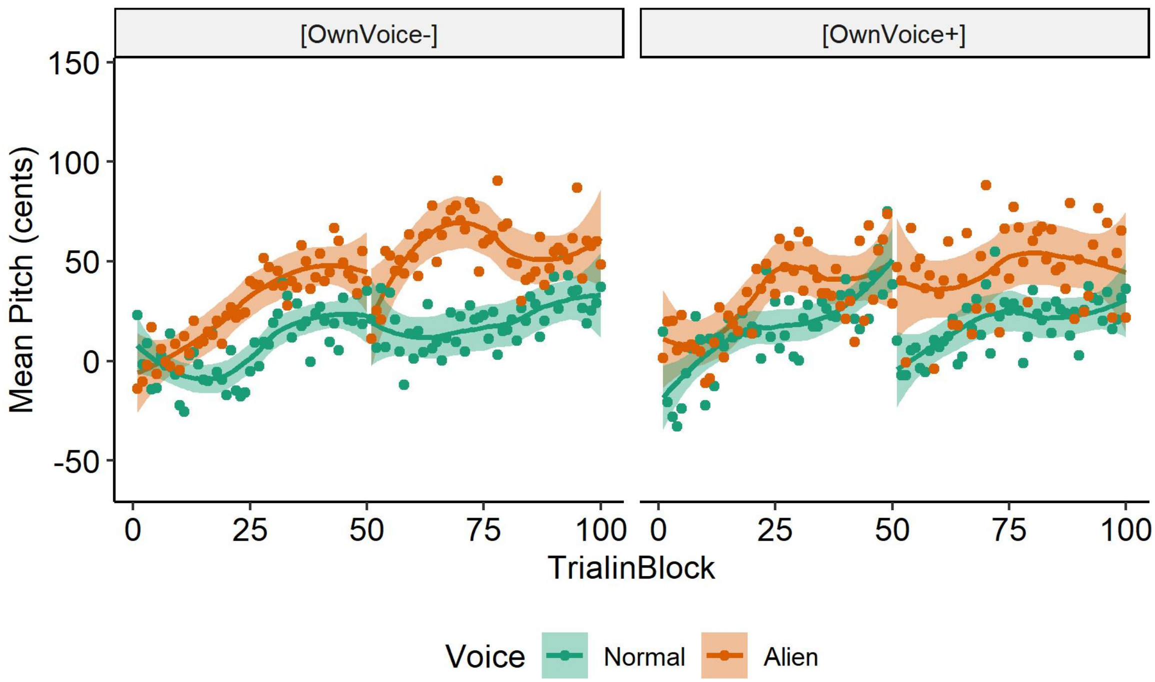

Figure 8. Change in mean pitch over the course of an experimental block as a function of feedback voice condition and participant group. The change in pitch is expressed as the change from the average pitch in the first 10 trials of a block. Participants got a break after 50 trials in each block.

Discussion

Overall, speakers compensated for unexpected pitch shifts in auditory feedback, both in the normal voice as well as in the alien voice feedback. This indicates that the sense of agency is flexible, and that speakers accepted the alien voice as a reference signal, by compensating for pitch deviations from it. This is in line with predictions from the inferential model, but not with the comparator model's perspective. It is interesting to highlight that speakers compensated even for downward pitch shifts in the alien voice condition. As the alien voice was a constant +500 cents pitch increase, downward pitch shifts in this condition brought the pitch closer to participants' actual production. Still, even in these cases participants tended to compensate by increasing pitch in response to an unexpected pitch decrease. In addition, we replicated earlier results showing that compensatory responses to large 400 cents pitch shifts were smaller and started later compared to small 100 cents pitch shifts (Korzyukov et al., 2017; Scheerer et al., 2013; Subramaniam et al., 2018).

In addition to the analysis of compensatory responses as an implicit index of sense of agency, explicit measures of the sense of agency were taken by means of a survey. Analysis of the participants' ratings revealed that the alien voice led to a decrease in the explicit sense of agency. Interestingly, there was considerable individual variability in how participants consciously experienced the alien voice feedback. A number of participants did not show a difference in their ratings between the normal voice and the alien voice block with respect to how similar the feedback was to their own voice. This result provides additional evidence that the sense of agency is very flexible in accommodating even a voice that was pitch-shifted by 500 cents. Next, the participants were into two groups depending on whether they rated the alien voice as less like their own voice compared to the normal voice or not. An exploratory factor analysis supported the use of this dimension to split the participants in two groups. We found an interaction between group and voice, suggesting that there was individual variability in the effect of the alien voice on feedback processing as a function of whether participants felt that the alien voice sounded less like their own voice. We want to stress that this result should be interpreted with caution, as it was based on a post-hoc group split, and resulted in an only barely significant interaction. Nevertheless, closer examination revealed that speakers who deemed the alien voice less like their own voice, showed a smaller response to pitch shifts in the alien voice compared to the other group. This result, by itself, is in line with the comparator model's predictions.

The alien voice block could have led – in these participants – to a loss of sense of agency and thus a reduction in compensatory responses. However, in contrast with the comparator model, a non-trivial number of participants did not show a difference in the sense of agency between the normal and the alien voice, neither in the implicit nor the explicit measures.

There was only a limited difference in responses to unexpected pitch shifts as a function of feedback voice, but the alien voice did lead to pitch alignment. Note the difference with the responses to short unexpected pitch shifts: these lead to clear compensation responses, whereas the large constant pitch shift led to pitch alignment over the course of the block. This is in line with previous studies that show speech alignment with voices that induced a sense of agency (Tajadura-Jiménez et al., 2017; Zheng et al., 2011). Participants in Zheng et al. (2011) showed a relative dissociation between pitch alignment and conscious perception of action agency: whereas mistimed feedback decreased introspective ratings of agency, participants showed no change in pitch alignment. On the other hand, in the current study, participants with a lower sense of agency with the alien voice as measured by their subjective rating, tended to show clearer pitch alignment with the alien voice. Additional research will be necessary to confirm this trend and to determine the causes of pitch alignment in these contexts.

Although there was individual variability in speakers' compensation responses and the effect of the alien voice as a function of the rating results, we observed few differences in feedback processing as a function of feedback voice type. Even in the [OwnVoice-] participants, who showed smaller compensation responses in the alien voice, this reduction was small and they still clearly compensated for the unexpected brief pitch shifts. This suggests that overall, the sense of agency was flexible and that speakers accepted the high-pitched alien voice as a new pitch referent. One possible explanation is that speakers adapted their forward model because the voice types were blocked in the current experiment. Previous studies show that prolonged exposure to consistently altered feedback (e.g., a constant pitch shift) leads participants to update their forward model (Houde & Jordan, 1998; Jones & Munhall, 2000).

As participants were exposed to the alien voice for a block of 100 consecutive trials, they may have had the time to learn to predict the novel voice sound (although in contrast to previous speech adaptation studies, this did not show up as compensatory change in production over the course of the block). In other words, after a while participants may have learned to predict the pitch increase in the alien voice, and therefore there was no longer a mismatch between predicted and observed auditory feedback in the alien voice condition. If so, this result could still be in line with the comparator model, suggesting that speakers may have had the time to adapt their forward model to match the alien voice's pitch shift, thus explaining the similar compensatory responses in both the normal and the alien voice blocks. Therefore, Experiment 2 served as a conceptual replication of the findings in Experiment 1. In addition, we investigated the role of consistent exposure to the new voice by randomizing voice type within each experimental block.

Experiment 2

Methods

All methods and materials were the same as in Experiment 1, except for the following.

Participants

Thirty-eight further participants (7 males, 31 females; average age = 21; SD = 3.0) took part in Experiment 2 in exchange for either course credit or payment. All participants were native speakers of Dutch and had no history of speech or hearing impairment. The experiment was approved by the local ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University. All participants signed a consent form in accordance with the declaration of Helsinki.

Experimental procedure

The procedure was the same as in Experiment 1, except for the voice type randomization. This time, both experimental blocks included trials with normal voice feedback and trials with alien voice feedback. In each block, participants received normal feedback during half of the trials, and alien voice feedback during the other half of the trials. The voice types were pseudo-randomized such that every quadruplet of four consecutive trials contained two normal voice and two alien voice trials.

After the experiment, we did not have the same statements as in Experiment 1 for participants to rate, because participants could not easily be told to rate statements on either normal voice or alien voice trials as these were mixed within experimental blocks. Instead, participants were asked open-ended questions about what they heard via the headphones, how the voice in the headphones sounded, and how it differed from their own voice. The goal of these open-ended questions was to probe participants to spontaneously mention whether they noticed the difference between the two voice types.

Analysis

The preprocessing steps were the same as in Experiment 1. The data of two participants were discarded because after preprocessing, there were fewer than 20 trials left for some conditions. For the remaining participants, on average 522.1 trials per participant were left (87.0%).

Results

Figure 9 shows the average pitch response contours in Experiment 2. Again, a cluster-based permutation test was run for each condition to compare the compensation response to the response contour to the corresponding null shift. Table 7 shows that in all but one condition, there was a significant compensation response. Only the response to the -400 cents shift in the alien voice condition did not differ from the response to a null shift. Upon visual inspection, this seemed driven by a brief decrease in the compensation response around 200-300ms, after which participants compensated again. Overall, these results suggest a flexible sense of agency, given that even in the alien voice condition, speakers tended to compensate for unexpected pitch shifts in almost all conditions, in line with the results of Experiment 1. These results are in line with predictions from the inferential model, but not from the comparator model's perspective. In line with the results in Experiment 1, also here responses to 400 cents pitch shifts seemed to start later compared to responses to 100 cents shifts.

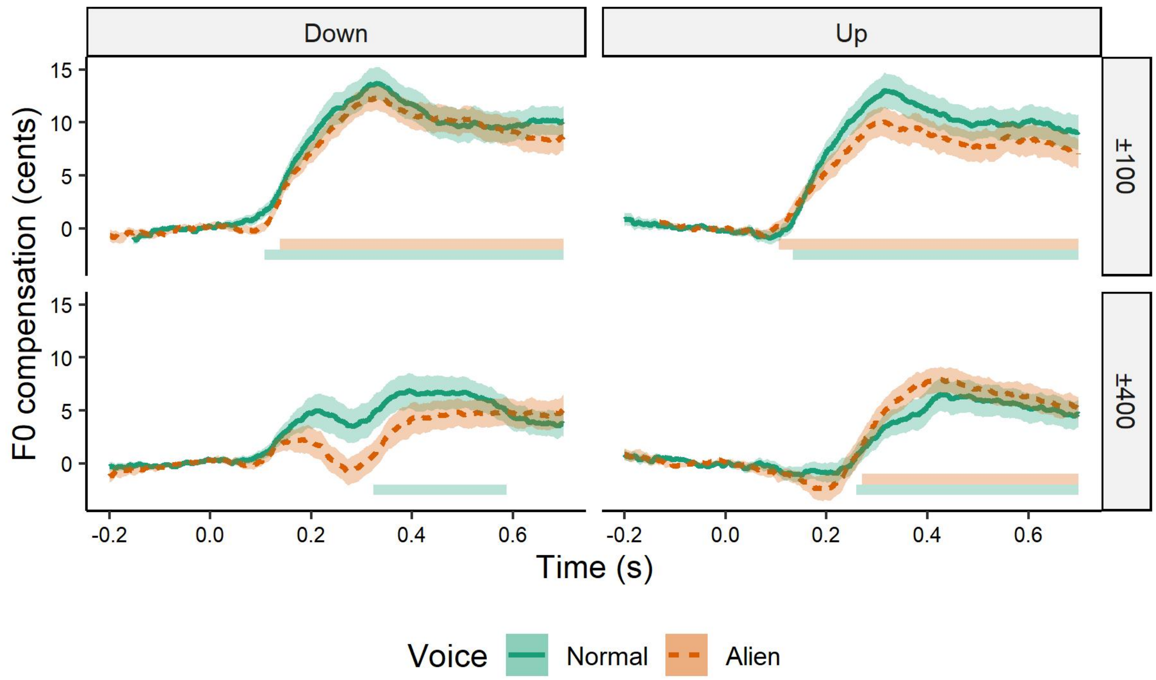

Figure 9. Average pitch compensation contours as a function of feedback voice, perturbation direction, and perturbation magnitude. Horizontal bars at the bottom of each panel indicate the extent of the largest cluster driving the difference between this condition and the corresponding zero shift perturbation.

Table 7. Results of cluster-based permutation tests in Experiment 2

Voice	Pitch Shift (cents)	Onset largest cluster (ms)	p	Cohen's d
Normal	100	133	< .001*	1.56
Normal	-100	107	< .001*	-1.33

Normal	400	259	< .001*	1.015
Normal	-400	323	.024*	-0.40
Alien	100	105	< .001*	1.39
Alien	-100	138	< .001*	-1.35
Alien	400	271	< .001*	1.13
Alien	-400		.28	-0.36

Next, the response magnitude was modeled as a function of voice type, shift direction, and shift magnitude in a linear mixed effects model (Figure 10). The fixed effects are shown in Table 8. There was a main effect of voice and shift magnitude, and trends for a two-way interaction between voice and shift magnitude, and a 3-way interaction between voice, shift magnitude, and shift direction. None of the other terms were close to significant. Thus, in contrast to Experiment 1, the overall response magnitudes in Experiment 2 were affected by voice type, with larger responses with normal feedback than with alien voice feedback (est. = 1.61, $t(35) = 2.60$, $p = .014$). The main effect of shift magnitude showed that participants compensated more for ±100 cents shifts than for ±400 cents shifts (est. = 4.71, $t(35) = 5.32$, $p < .001$). A closer analysis of the three-way interaction term showed that for the downward pitch shifts, the effect of feedback voice was similar for small and large pitch shifts, showing a decrease of about 2 to 3 cents (although not significant, -100 cents shift: est. = 2.37, $t(42.9) = 1.94$, $p = .059$; -400 cents shift: est. = 1.89, $t(40.5) = 1.31$, $p = .20$). However, for the upward shifts, the effect was similar for small shifts (est. = 2.88, $t(42.1) = 2.24$, $p = .030$), but absent for large shifts (est. = -0.69, $t(42.5) = -0.549$, $p = .59$). So it seems the three-way interaction was mainly driven by an absence of the overall alien voice-related decrease in compensation for the +400 cents shifts.

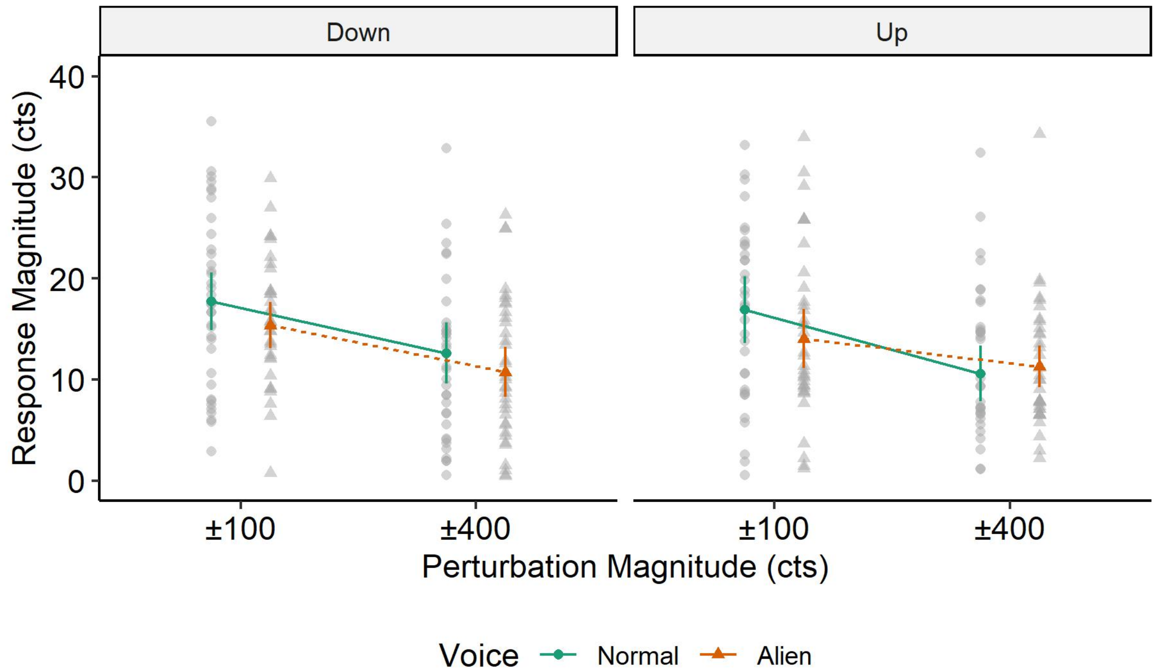

Figure 10. Response magnitude as a function of perturbation magnitude, feedback voice, and perturbation direction.

Table 8. Fixed effects of the model Response Magnitude ~ Voice*ShiftMagnitude*ShiftDirection + (pertMag * pertDir + Voice + Voice:pertMag + Voice:pertDir | Participants)

	SS	df	F	p
Voice	**74.11**	**1, 35**	**6.75**	**.014***
ShiftMagnitude	**310.81**	**1, 35**	**28.30**	**< .001***
ShiftDirection	4.00	1, 35	0.36	.55
Voice:ShiftMagnitude	42.85	1, 35	3.90	.056
Voice:ShiftDirection	3.29	1, 35	0.30	.59
ShiftMagnitude:ShiftDirection	0.44	1, 35	0.040	.84

Voice:ShiftMagnitude:ShiftDirection	42.58	1, 35	3.88	.057

Questionnaire Results

As we couldn't have participants rate statements specifically on the alien voice and normal voice trials separately (given that they were intermixed within a block), the participants were asked open questions about how the feedback sounded to them. Participants described the alien voice as much higher than their own voice. Some participants mentioned it sounded like a voice on helium, robot-like, or like a small child's voice.

Discussion

Similar to the results of Experiment 1, the participants in Experiment 2 showed stronger responses (which started earlier) to smaller pitch shifts than to the larger pitch shifts. But different from Experiment 1, there was now an overall tendency to make smaller responses to pitch shifts in the alien voice feedback compared to the normal voice feedback. It is possible that the randomization of voice may have led to a decrease in responses to alien voice pitch shifts. This is in line with the comparator model's prediction that the alien voice would evoke less compensation. The contrast with the results of Experiment 1 could indicate that, in Experiment 1, the blocked design allowed participants to predict the alien voice. Another possibility is that in Experiment 2, there was a higher proportion of participants who subjectively felt a difference in sense of agency between the normal and the alien voice (i.e., similar to the [OwnVoice-] group in Experiment 1). Because of the randomized design, we could not have the same survey in Experiment 2, but responses to the open questions suggest that the difference between normal and alien voice was subjectively more salient in Experiment 2.

In addition, there was a trend of a three-way interaction between voice, perturbation magnitude, and perturbation direction. This interaction suggested that the effect of perturbation magnitude was smaller for upward shifts in the alien voice feedback compared to other conditions. While a smaller effect of perturbation magnitude in alien voice feedback was hypothesized by the comparator model, it is not clear why this should depend on perturbation direction. Another way to look at this, is to consider that the three-way interaction shows a quite different effect of feedback voice in the +400 cents shifts compared to all other shifts. While the alien voice leads to a decrease in response magnitude of about the same magnitude in most conditions (ranging from a 1.89 to a 2.88 cents decrease), in the +400 cents shifts, alien voice leads to no change (numerically a 0.69 cents increase). A potential factor may be the interaction between the constant pitch shift of the alien voice (a +500 cents shift) and the short +400 cents pitch shift, leading to a temporary 900 cents difference with the produced pitch. It is possible that acceptance of the alien voice as a new reference, as evidenced by the compensations in almost all conditions, is not complete, and both absolute pitch difference between the produced pitch and the feedback pitch as well as the unexpected sudden change in the feedback signal may affect subsequent vocalization responses.

General Discussion

Overall, Experiments 1 and 2 show only a limited effect of feedback voice. In Experiment 1, participants' responses to unexpectedly pitch-shifted feedback overall did not vary as a function of feedback voice, but there was some individual variability. Participants who showed a weaker sense of agency over the alien feedback in their subjective ratings tended to show smaller responses to the alien feedback compared to participants who showed no such effect in their ratings. Experiment 2 served as a conceptual replication of the findings in Experiment 1. In addition, by randomizing condition order, we were able to examine whether the results hold when the feedback voice was unpredictable. There was an overall small effect of smaller responses in the alien voice compared to the normal voice condition, but participants still showed clear compensatory responses in almost all conditions. These results suggest that our alien voice manipulation did not strongly affect the participants' sense of agency over the auditory feedback, even though this manipulation radically changed the way their speech sounded. In contrast, speakers seemed to accept the high-pitched alien voice quickly as a new referent, against which compensatory responses to unexpected pitch shifts were given. This was especially clear in the case of unexpected pitch decreases in the alien voice conditions. In these cases, the pitch shift is a pitch decrease, while the resulting pitch is still higher than the actually produced pitch. Still, participants tended to compensate for the pitch decrease by increasing their pitch. These results are in line with earlier studies that point to a flexible sense of vocal agency, such as speakers accepting another person's voice as their own (Zheng et al., 2011) or accepting a lexical error in the feedback as their own (Lind et al., 2014).

These results are in line with the predictions by the inferential model of the sense of agency. Participants compensated for unexpected brief pitch shifts, even when the auditory feedback sounds alien. If we take compensatory responses as indicative of the presence of a sense of vocal agency, this suggests that a match between predicted feedback and observed feedback is not crucial in generating the sense of agency. Other cues such as simultaneity between articulation and alien voice feedback, the absence of an alternative source of the auditory input, or correspondences in other acoustic parameters besides pitch (e.g., spectral characteristics) may have led participants to generate a sense of agency over the alien voice. As the inferential account does not deny a role for the comparison between predicted and observed auditory feedback (Moore et al., 2009; Synofzik et al., 2008), the individual variability with respect to the effect of the alien voice on compensatory responses is not necessarily in contrast with this model. It may be the case that the weight assigned to each of the various cues that together lead to the sense of agency varies across speakers. In the case of Experiment 1, the [OwnVoice-] participants may assign more weight to the feedback comparison in determining agency, compared to the [OwnVoice+] participants. In Experiment 2, the effect of the alien voice suggests that the comparison between predicted and observed feedback plays a role in the sense of agency. This is in line with the view that the sense of agency is the result of a multi-layered, dynamic process (Moore, 2016; Synofzik et al., 2008). In addition, the randomized design in Experiment 2 may have made the alien voice manipulation more salient, which in turn could contribute to a lower probability of inferring a sense of agency.

Alternatively, the current results could be interpreted within the comparator model framework. The [OwnVoice-] participants in Experiment 1, and the participants in Experiment 2 show that the alien voice led to decreased compensatory responses to unexpected brief pitch shifts, which we take to be an implicit index of a lower sense of agency in this condition. This result is in line with the prediction by the comparator model. The alien voice does not match with the participants' expected pitch, and therefore signals that the auditory input might not be self-generated. In addition, the rating results for the [OwnVoice-] participants in Experiment 1 also serve as an explicit indication that participants had a lower sense of agency. Although the fact that a non-trivial number of participants in Experiment 1 did not show a reduction of agency in the explicit nor the implicit measures is in contrast with the comparator model's predictions, this could be accounted for if participants learned to predict the sound of the alien voice over the course of the experiment. This is in line with previous studies showing that prolonged exposure to manipulated feedback leads participants to adapt their internal forward models (Houde & Jordan, 1998; Jones & Munhall, 2000). In the current study, however, if there was adaptation, it was not associated with compensatory changes in speech production typically seen. Experiment 2 showed that even when trials were randomized, participants compensated for the brief unexpected pitch shifts in almost all conditions. If this is due to forward model adaptation, this suggests that participants were able to adapt their predictions within the first 600 to 800ms of the trial (the time between speech onset and the first unexpected pitch shift). It may have helped participants that even in the randomized design, there were only two feedback voice conditions. In other words, even if the occurrence of alien voice trials was unpredictable, once detected, the pitch manipulation in these trials was always the same (always a +500 cents difference). This way, participants may get familiarized with the alien voice manipulation and thus quickly adapt their feedback predictions once it is detected at speech onset.

Further research will be necessary to disentangle these accounts of the current results and of the sense of agency in speech control. Note, as stated in the introduction, that a comparator view and an inferential view of the sense of agency are not necessarily mutually exclusive, as suggested by the synthesis presented by Moore et al. (2009). Both the comparison between predicted and perceived sensory feedback as well as external contextual cues could play a role in determining the sense of agency in speech. A potential avenue for further studies should entail further characterization of the individual variability in the vocal sense of agency as shown here in Experiment 1, and how it is handled by the different theoretical frameworks.

An interesting finding, tangential to the preceding discussion, is that the current study suggests that participants responded to the change in the feedback pitch, rather than to the absolute difference in pitch between what was actually produced and what was perceived. Because the alien voice entailed a constant pitch shift of +500 cents, the short pitch decreases led to feedback that was still higher than participants' actual pitch, even for downward pitch shifts. Still, participants responded by increasing their pitch in response to an unexpected pitch decrease. This suggests that speakers were compensating for the sudden decrease in pitch from the alien voice baseline, rather than for the absolute pitch difference between the feedback and their actual production. It is possible therefore that the sense of agency is driven by temporal covariance, rather than the absolute difference between predicted and observed feedback. In other words, even in the alien voice, small fluctuations in pitch production were still reflected in the feedback (except for a constant 500 cents baseline shift), leading to temporal covariance between the articulation and the feedback. This idea is in line with the principle of consistency in Wegner and Wheatley (1999)'s theory of apparent mental causation, which states that the sense of agency requires that sensory feedback is consistent with a prior intention. In addition, a recent study showed a role for motor-visual regularities in the sense of agency over dots controlled by a computer mouse (Wen & Haggard, 2020). These authors suggest that the sense of agency may depend on both ongoing covariance and on detecting prediction errors in the comparison between predicted and perceived sensory feedback. Indeed, this hypothesis is not mutually exclusive with either the inferential or the comparator model accounts of the current results. With respect to the current study, this suggestion is relatively speculative. Further studies will be necessary to test this hypothesis.

In this context it is interesting to again refer to the literature on the so-called rubber-hand illusion (Botvinick & Cohen, 1998). In these studies, just as in the present study, there is covariance between the predicted sensory consequences of movement and the observed sensory feedback. The ’rubber’ hand’s movements differ from the participant hand’s movements only by a constant shift in spatial location, and potentially a number of constant visual characteristics of the hand. The movement path, including random fluctuations, errors, or idiosyncrasies are reflected in the ‘rubber’ hand’s movements and thus lead to covariance between the rubber hand’s movement and the participant hand’s movement. Interestingly, the rubber-hand illusion can also be generated if there is a clear mismatch between constant visual characteristics of the rubber hand and the participant’s hand, such as skin color (Maister et al., 2013). This suggests that our body representation demonstrates remarkable flexibility and is not constant over time (Banakou & Slater, 2014). The current results suggest that this may also apply to vocal agency: participants compensated for pitch shifts even in the high-pitched alien voice, suggesting the alien voice feedback was treated as their own voice, or at least as the result of their own articulation. This was the case both when participants were exposed to the alien voice consistently across trials (Experiment 1), or when the feedback voice type was unpredictable (Experiment 2), suggesting participants adapt to novel voice characteristics very quickly. In line with these results, an earlier study showed, using explicit measures through an agency survey, that adult participants even could show a sense of agency over a child's body and a child's voice in virtual reality (Tajadura-Jiménez et al., 2017).

In addition to responses to the unexpected brief pitch shifts, participants in Experiment 1 also aligned their pitch to the alien voice over the course of the block. This is reminiscent of previous studies where participants aligned their pitch to a stranger's voice (Zheng et al., 2011) or pitch-shifted feedback in a virtual environment (Tajadura-Jiménez et al., 2017). The authors in these studies suggest that this pitch alignment may be driven by the sense of agency, although it is not clear why a sense of agency over altered feedback would lead to pitch alignment. One could speculate that pitch alignment could be an indication of an attempt to reduce the mismatch between articulation and feedback. However, the individual variability shown in the current study casts doubt on the hypothesized causal link between sense of agency and pitch alignment. The current findings suggest that it is mostly the participants who show a decreased sense of agency in their subjective ratings, that show pitch alignment (although also participants in the [OwnVoice+] group showed pitch increase). One possible interpretation is that it is not the sense of agency, but simply the exposure to high-pitched auditory input that leads to pitch alignment, similar to speech alignment (also termed phonetic convergence) with an interlocutor during conversation (Pardo, 2006). A more salient pitch difference, as might be the case for the [OwnVoice-] group, could lead to more alignment (although this is not necessarily the case in conversational phonetic convergence). An alternative interpretation is that given that sense of agency could exist on a continuum ranging from complete sense of agency to none at all, there may be an inverted U-shaped relationship between the sense of agency and pitch alignment. It may be that the sense of agency causes pitch alignment when there is some, but not complete, sense of agency. With complete sense of agency, one could imagine it serves no purpose for the speaker to change articulation to match their own feedback, given that the feedback is the result of their previous articulation already. Future research should try to disentangle these alternative interpretations.

Finally, the current results showed that responses to the larger pitch shifts (±400 cents) started later compared to responses to smaller shifts (±100 cents). This concurs at least with what can descriptively be seen on the figures in some previous studies (Korzyukov et al., 2017; Subramaniam et al., 2018). These studies have argued that the larger shifts lead to a reduced response because of a diminished sense of agency. The same reason could possibly also explain the later response onset. We speculate that perhaps a reduced sense of agency leads to a weaker error signal, and thus not only to smaller but also slower responses. Currently, it is unclear what drives this difference in response latency. Further research should further investigate this issue.

In conclusion, the current results show that the sense of agency is highly flexible in speech production and may derive from temporal covariance between expected and observed sensory feedback, rather than an absolute match. These results raise questions about the content of sensory predictions and the internal representation of one's own voice. With respect to sensory predictions, temporal variation may be more prominently represented than, for example, absolute pitch. This suggests that the forward model does not represent (only) instantaneous pitch, but also contains a time dimension. Internal representations of one's own voice are more flexible than often assumed, in line with recent studies beyond speech, showing that one's representation of one's body is flexible.

Acknowledgements

We are grateful to Pascal Mestdagh for his help with acoustic measurements and experimental setup. M.K.F. was supported by the Special Research Fund (BOF) at Ghent University. A.L. was supported by the Swedish Research Council (Grant No. 2015-00446) and the Crafoord Foundation (Grant No. 20170734), L.H. and P.J. were supported by the Swedish Research Council (Grant No. 2014-1371) and the Crafoord Foundation.

References

Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, *111*(49), 17678–17683. https://doi.org/10.1073/pnas.1414936111

Banakou, D., & Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Scientific Reports*, *7*(1), 14227. https://doi.org/10.1038/s41598-017-14620-5

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2014). Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain and Cognition*, *84*(1), 97–108. https://doi.org/10.1016/j.bandc.2013.11.007

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America*, *132*(4), 2468–2477. https://doi.org/10.1121/1.4746984

Blakemore, Sarah-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. https://doi.org/10.1038/2870

Blakemore, Sarah-Jayne, & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology*, *13*(2), 219–224. https://doi.org/10.1016/S0959-4388(03)00043-6

Blakemore, Sarah-Jayne, Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237–242. https://doi.org/10.1016/S1364-6613(02)01907-1

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer [Computer Program]* (6.0.33).

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. https://doi.org/10.1038/35784

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Caspar, E. A., Cleeremans, A., & Haggard, P. (2015). The relationship between human agency and embodiment. *Consciousness and Cognition*, *33*, 226–236. https://doi.org/10.1016/j.concog.2015.01.007

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F[sub 0] responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Dennett, D. Clement. (1991). *Consciousness explained*. Penguin Books.

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*(4), 1458–1467. https://doi.org/10.3758/s13423-018-1494-x

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., Wartenberg, T., & Lind, A. (2019). Does passive sound attenuation affect responses to pitch-shifted auditory feedback? *The Journal of the Acoustical Society of America*, *146*(6), 4108–4121. https://doi.org/10.1121/1.5134449

Frith, C. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition*, *14*(4), 752–770. https://doi.org/10.1016/J.CONCOG.2005.04.002

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14

Hawco, C. S., Jones, J. A., Ferretti, T. R., & Keough, D. (2009). ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*, *46*(6), 1216–1225. https://doi.org/10.1111/j.1469-8986.2009.00875.x

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*(82). https://doi.org/10.3389/fnhum.2011.00082

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246. https://doi.org/10.1121/1.1288414

Kalckert, A., & Ehrsson, H. H. (2012). Moving a Rubber Hand that Feels Like Your Own: A Dissociation of Ownership and Agency. *Frontiers in Human Neuroscience*, *6*, 40. https://doi.org/10.3389/fnhum.2012.00040

Korzyukov, O., Bronder, A., Lee, Y., Patel, S., & Larson, C. R. (2017). Bioelectrical brain effects of one's own voice identification in pitch of voice auditory feedback. *Neuropsychologia*, *101*, 106–114. https://doi.org/10.1016/j.neuropsychologia.2017.04.035

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *LmerTest: Tests in Linear Mixed Effects Models. R Package version 2.0-33* (2.0-33).

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F[sub 0] responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*(6), 2845. https://doi.org/10.1121/1.1417527

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' Acceptance of Real-Time Speech Exchange Indicates That We Use Auditory Feedback to Specify the Meaning of What We Say. *Psychological Science*, *25*(6), 1198–1205. https://doi.org/10.1177/0956797614529797

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *122*(5), 951–957. https://doi.org/10.1016/j.clinph.2010.08.010

Lush, P. (2020). Demand Characteristics Confound the Rubber Hand Illusion. *Collabra: Psychology*, *6*(1), 22. https://doi.org/10.1525/collabra.325

Maister, L., Sebanz, N., Knoblich, G., & Tsakiris, M. (2013). Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition*, *128*(2), 170–178. https://doi.org/10.1016/J.COGNITION.2013.04.002

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/DOI 10.1016/j.jneumeth.2007.03.024

Moore, J. W. (2016). What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01272

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Puckette, M. (1996). Pure data. *Proceedings of the International Computer Music Conference*, 269–272.

R Core Team. (2018). *R: a language and environment for statistical computing*. R Foundation for Statictical Computing.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, *271*(1–2), 16–25. https://doi.org/10.1016/J.HEARES.2010.09.001

Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: Implications for sense of agency and its disorders. *Consciousness and Cognition*, *37*, 1–7. https://doi.org/10.1016/j.concog.2015.07.011

Scheerer, N. E., Behich, J., Liu, H., & Jones, J. A. (2013). ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience*, *240*, 176–185. https://doi.org/10.1016/j.neuroscience.2013.02.054

Scheerer, N. E., & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *The European Journal of Neuroscience*. https://doi.org/10.1111/ejn.12734

Shadmehr, R., & Mussa-Ivaldi, F. a. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, *14*(5), 3208–3224. https://doi.org/8182467

Singmann, H., Bolker, B., Westfall, J., Frederik, A., & Ben-Shachar, M. S. (2019). *afex: Analysis of Factorial Experiments* (R package version 0.25-1).

Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S. S., & Houde, J. F. (2018). Reality Monitoring and Feedback Control of Speech Production Are Related Through Self-Agency. *Frontiers in Human Neuroscience*, *12*, 82. https://doi.org/10.3389/fnhum.2018.00082

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a Child-Like Talking Virtual Body Influences Object Size Perception, Self-Identification, and Subsequent Real Speaking. *Scientific Reports*, *7*(1), 9637. https://doi.org/10.1038/s41598-017-09497-3

Tsakiris, M., & Haggard, P. (2005). The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-Attribution. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 80–91. https://doi.org/10.1037/0096-1523.31.1.80

Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition*, *15*(2), 423–432. https://doi.org/10.1016/J.CONCOG.2005.09.004

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480–492. https://doi.org/10.1037/0003-066X.54.7.480

Wen, W., & Haggard, P. (2020). Prediction error and regularity detection underlie two dissociable mechanisms for computing the sense of agency. *Cognition*, *195*, 104074. https://doi.org/10.1016/j.cognition.2019.104074

Wolpert, D., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat Neurosci*, *3*(Suppl), 1212–1217.

Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a Stranger's Voice as Being One's Own: A 'Rubber Voice' Illusion? *PLoS ONE*, *6*(4), e18655. https://doi.org/10.1371/journal.pone.0018655