Drifting pitch awareness after exposure to altered auditory feedback

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

[1]Department of Experimental Psychology, Ghent University, Ghent, Belgium

[2]Lund University Cognitive Science, Lund University, Lund, Sweden

Corresponding author: Matthias K. Franken, Department of Psychology, McGill University, Room 718, 2001 McGill College Avenue, Montreal QC H3A 1G1, Canada, phone: +1-514-398-6111, e-mail: matthias.franken@mcgill.ca

*Currently at Department of Psychology, McGill University, Montreal, Canada

Abstract

Various studies have claimed that the sense of agency is based on a comparison between an internal estimate of an action's outcome and sensory feedback. With respect to speech, this presumes that speakers have a stable pre-articulatory representation of their own speech. However, recent research suggests that the sense of agency is flexible and thus in some contexts we may feel like we produced speech that was not actually produced by us. The current study tested whether the estimated pitch of one's articulation (termed 'pitch awareness') is affected by manipulated auditory feedback. In four experiments, fifty-six participants produced isolated vowels while being exposed to pitch-shifted auditory feedback. After every vocalization, participants indicated whether they thought the feedback was higher or lower than their actual production. After exposure to a block of high-pitched auditory feedback (+500 cents pitch shift), participants were more likely to label subsequent auditory feedback as "lower than my actual production", suggesting that prolonged exposure to high-pitched auditory feedback led to a drift in participants' pitch awareness. The opposite pattern was found after exposure to a constant -500 cents pitch shift. This suggests that pitch awareness is not solely based on a pre-articulatory representation of intended speech or on a sensory prediction, but also on sensory feedback. We propose that this drift in pitch awareness could be indicative of a sense of agency over the pitch-shifted auditory feedback in the exposure block. If so, this suggests that the sense of agency over vocal output is flexible.

Introduction

Sense of agency refers to the feeling that one is in control of one's actions (Haggard, 2017). A core aspect of the sense of agency is the association between an action and its sensory outcome. Various researchers have previously suggested that the generation of the sense of agency is based on a comparison between an internal estimate of an action's outcome and the observed sensory feedback. If there is no discrepancy, a sense of agency is generated (Blakemore et al., 2002; Frith, 2005; Wolpert & Ghahramani, 2000). Applied to the case of speech production, this presumes that speakers have a stable pre-articulatory representation of their own speech, to serve as a benchmark against which incoming sensory signals can be compared. An alternative account suggests that the production process itself does not give rise to such detailed and fixed representations and that the sense of agency is instead *inferred* from various sources of information, including sensory feedback (Lind et al., 2014a). Thus, a stable representation of speech may only be available 'after the fact'. The current study investigates these hypotheses by focusing on whether exposure to altered auditory feedback carries over to affect speakers' representation of their own speech output in subsequent utterances. If it does, this is problematic for an account where comparing this representation to auditory feedback is the sole contributing factor to the sense of agency. Rather, the data would be more in line with inferential accounts, which hold that the aforementioned comparison process is just one of multiple factors contributing to the sense of agency.

Recently, a number of studies have suggested that the sense of agency over speech is flexible. Zheng et al. (2011) replaced speakers' auditory feedback with a stranger's voice. The participants accepted the stranger's voice as their own, showing a sense of agency over the auditory input. At the lexical level, Lind et al. (2014b) had participants carry out a Stroop task, while auditory feedback was replaced such that participants said one thing but heard themselves saying something else. Participants often accepted

the inserted feedback as self-produced, and thus accepted that they made an error, while in reality they did not. Recently, we showed that participants corrected for unexpected pitch shifts in auditory feedback regardless of whether the feedback sounded like their own voice or not (Franken et al., 2021), suggesting that speakers had a sense of agency even over a high-pitched 'alien' voice that sounded unlike their own. Together, these studies indicate that the sense of vocal agency is flexible, and that speakers' representation of their vocal output is not (only) based on a pre-articulatory representation of vocal output, but also on auditory feedback.

Similarly, in non-speech motor control, studies have suggested that internal representations of the sensory consequences of our own movements are quite flexible. In the rubber hand illusion, participants experience the illusion that a rubber hand is part of their body (Botvinick & Cohen, 1998). The illusion can occur through simultaneous visuotactile stimulation of the rubber hand and the participant's actual hand, or through simultaneous movement of both hands. Interestingly, several studies have shown that this illusion is accompanied by so-called proprioceptive drift: the perceived position of the participant's own hand drifts towards the location of the rubber hand (Botvinick & Cohen, 1998; Tsakiris et al., 2006; Tsakiris & Haggard, 2005). This suggests that the perception of the current location of one's own body parts, or one's body awareness, is affected by an integration of multimodal sensory information. Based on these studies, it has been argued that proprioceptive drift is associated with the sense of agency and can be used as a quantitative index of the rubber hand illusion (but see Lush (2020) for a discussion of demand characteristics in this literature).

Another phenomenon that has been associated with a sense of agency over a voice is the pitch alignment speakers show with the heard voice in a number of studies. Zheng et al. (2011) showed that replacing a speaker's auditory feedback with a high-pitched voice of another speaker led to pitch alignment: speakers tended to shift their pitch towards the pitch of the heard voice. This did not,

however, correlate with a subjective measure of agency. Two more recent studies show that the same effect is observed in cases where auditory feedback is not replaced by another speaker, but is pitch-shifted (Franken et al., 2021; Tajadura-Jiménez et al., 2017). In both cases, speakers showed a sense of agency over the manipulated feedback and tended to align pitch with the pitch-shifted feedback. Note that the magnitude of the difference between the pitch produced by the participant and the feedback's pitch in these studies is larger than typically used in the vocal control literature, where pitch-shifted feedback usually leads to compensation, rather than alignment (Jones & Munhall, 2000). A study that made use of immersion in a virtual environment showed that actual articulation by the participant is not even necessary to generate a sense of agency over auditory input, while participants still showed a pitch alignment effect in subsequent speech production (Banakou & Slater, 2014). However, this was not the case when immersion in the virtual body was evoked by visuotactile stimulation only, as opposed to visuomotor synchronous experience (Banakou & Slater, 2017). Together, these studies suggest that a sense of agency may be associated with pitch alignment to the heard voice, although the nature of this association is still unclear.

The current study aims to investigate the flexibility of a speaker's estimate of their own vocal output. In their study, Tsakiris et al. (2006) define the concept of body awareness as "the conscious experience of the location of a specific body-part in space". Here, we examine speaker's awareness of an aspect of their own speech output by defining what can be considered an equivalent concept in auditory space, which we call *pitch awareness*. We take this term to refer to the conscious experience of the pitch of one's own vocalizations. It has previously been established that prolonged exposure to altered auditory feedback leads to perceptual changes in addition to speech motor learning (Lametti et al., 2014; Shiller et al., 2009). These studies show that exposure to spectrally altered feedback leads to shifts in phoneme boundaries in a later speech perception task. In the current study, however, we focus on how pitch awareness is affected during speech production. If sense of agency over speech is flexible, exposure to

pitch-shifted auditory feedback should lead to a drift in pitch awareness. Specifically, we expect that

exposure to high- or low-pitched auditory feedback, where pitch is either increased or decreased by 500

cents, will affect speakers' pitch awareness such that their vocalizations seem higher (or lower) to them

than in reality. In contrast, from a comparator model's perspective, speakers' internal representation of

their own speech production needs to be stable in order to serve as a benchmark against which auditory

feedback is compared, so it should not be affected by pitch-shifted auditory feedback. In addition, while

participants in previous adaptation studies typically show compensation for small feedback

manipulations, we expect participants to follow the feedback manipulation, as typically seen with pitch

shifts of this magnitude, which is much larger than often used in altered auditory feedback experiments.

Finally, we will test whether changes in pitch awareness are indeed due to exposure to pitch-shifted

auditory feedback, and not simply due to exposure to high- or low-pitched auditory input in general

(which would not necessarily include a sense of agency).

Methods

All data and analysis scripts are available on the Open Science Framework, DOI 10.17605/OSF.IO/3KSQ2.

Participants

Fifty-six participants volunteered to take part in the study in exchange for course credit or a small

monetary reward. They signed an informed consent in accordance with the declaration of Helsinki. The

participants were randomly assigned to take part in one of four experiments (see Table 1). Fourteen of

them were assigned to take part in Experiment 1 (13 female and 1 male, mean age = 20.7 years, SD =

1.9), fourteen were assigned to Experiment 2 (11 female and 3 male, mean age = 21 years, SD = 1.46),

fourteen were assigned to Experiment 3 (10 female and 4 male, mean age = 24 years, SD = 3.34), and

fourteen were assigned to Experiment 4 (10 female and 4 male, mean age = 22.5 years, SD = 3.1). All

experimental procedures were approved by the ethical committee of the faculty of psychology and

educational sciences of Ghent University.

Table 1. Overview of the four experiments, varying by experimental task in the exposure block, and the direction of the pitch

shift. See Procedure for details.

	N	Exposure Block	Pitch shift
Experiment 1	14	Vowel production with pitch-shifted feedback	+500 cents[a]
Experiment 2	14	Vowel production with pitch-shifted feedback	-500 cents
Experiment 3	14	Passive listening to pitch-shifted recordings	+500 cents
Experiment 4	14	Passive listening to pitch-shifted recordings	-500 cents

[a] The unit "cents" refers to a psychometric scale to express a frequency interval, where doubling the frequency in Hertz

corresponds to a 1,200 cents increase. An interval of 100 cents reflects the interval between, for example, C and C# in western

music. See below (Methods section) for more information.

Procedure

In all experiments, participants wore headphones and were instructed to produce one of five Dutch

vowels (/i/, /e/, /a/, /o/, or /u/) as soon as a visual cue appears (Figure 1A). The visual cue was the Dutch

spelling for the corresponding vowel (<ie>, <ee>, <aa>, <oo> and <oe>, respectively). During vowel

production, participants received pitch-shifted auditory feedback through the headphones (Figure 1B).

Participants had 2s time to produce the vowel. The feedback was manipulated for the entire duration of

the trial. In blocks 1 (pretest) and 3 (posttest), the auditory feedback was shifted by 0, 30, 60, 100, or

150 cents, up or down, yielding trials with one out of ten possible pitch shifts (including the 'zero' pitch

shift twice). Every set of 10 consecutive trials contained all 10 pitch shifts, in random order. After every

trial, a question appeared on the screen, asking whether the participant thought the sound in the

headphones was lower or higher than their actual pitch. Participants were instructed to respond by

button press. They pressed the left arrow when they thought the pitch in the feedback was decreased,

and pressed the right arrow when they felt the pitch was increased. These response buttons

corresponded to the words 'lager' (Dutch for 'lower') and 'hoger' ('higher'), printed left and right on the

screen. The task and the design were the same for both the pretest and the post-test blocks across all

four experiments. Experiments differed only in the intervening exposure block. For Experiments 1 and 2,

in the exposure block, participants were instructed again to produce the five Dutch vowels on cue as in

the test blocks. However, the auditory feedback in all trials of the exposure block was manipulated by a

large constant 500 cents pitch shift, and no questions about the feedback were asked. The large pitch

shift was an increase in Experiment 1 (+500 cents), and a decrease in Experiment 2 (-500 cents). See

Figure 1B for an overview of altered feedback across Experiments 1 and 2. In contrast, in Experiments 3

and 4, participants were not asked to produce any vowels in the exposure block. Instead, they passively

listened to recordings of their own voice (recorded during a calibration phase that preceded the

experiment). All the recordings in the exposure block were pitch-shifted by +500 cents in Experiment 3,

and by -500 cents in Experiment 4. This way, the four experiments differed only during the exposure

block, and they differed in the participants' task (either active production with auditory feedback in

Experiments 1 and 2, or passive listening in Experiments 3 and 4), and in the direction of the pitch shift

(+500 cents in Experiments 1 and 3, and -500 cents in Experiments 2 and 4), see Table 1. While our main

interest was in whether altered auditory feedback affects pitch awareness, and thus in the effects in

Experiments 1 and 2, Experiments 3 and 4 served to rule out that effects were due to mere perceptual

exposure to high- or low-pitched auditory input. In all four experiments, the pretest contained 100 trials,

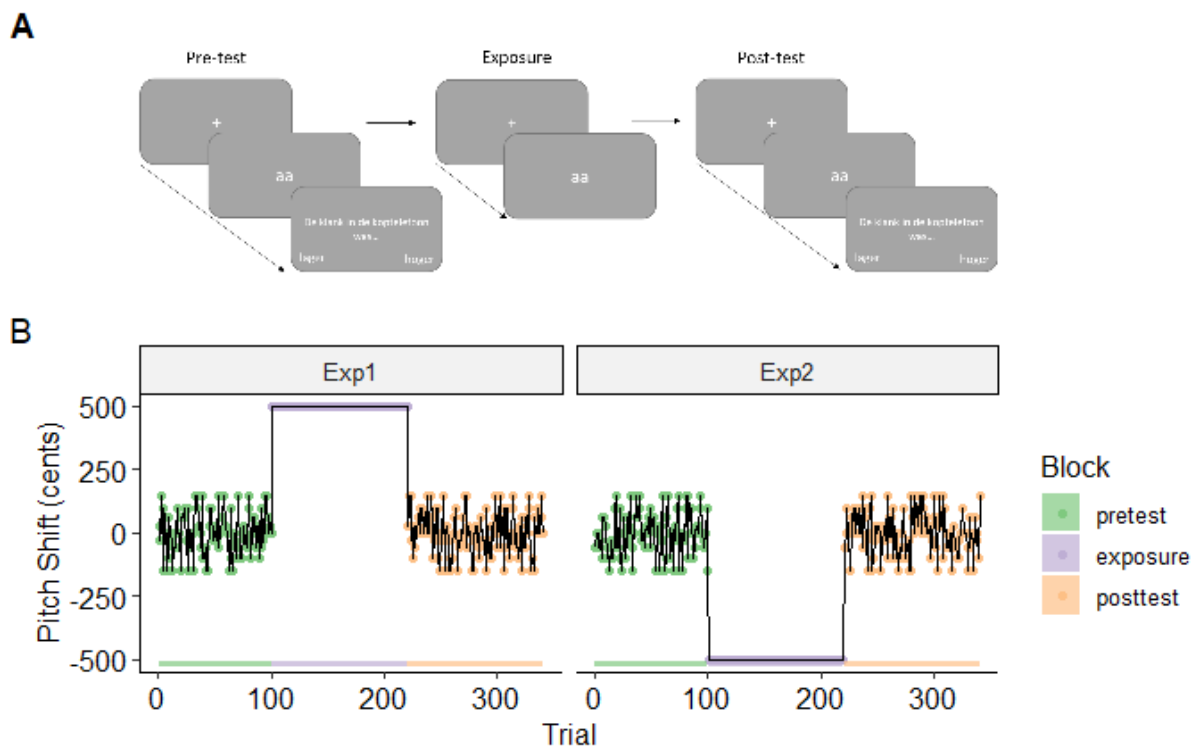the exposure block and the posttest each 120 trials.

Figure 1. Experimental design. (A) Illustration of visual cues and sequence of the experimental design. These visual cues were the same in all four experiments. After a fixation cross, a visual cue indicated which vowel participants were to utter. In the pre-test and post-test blocks, this was followed by a question, asking participants whether they felt the sound in the headphones was higher or lower than their actual pitch. In Experiments 3 and 4, although participants did not need to vocalize in the exposure block, they still were exposed to the visually presented vowels. (B) Example illustration of auditory feedback pitch shifts in Experiment 1 (left) and Experiment 2 (right), with the pretest and posttest blocks both including up- and down shifts of varying magnitudes (same in both experiments) and the exposure blocks including constant 500 cent shifts up (Experiment 1) or down (Experiment 2).

After the experiment, participants filled out a questionnaire that asked about their subjective experience of the experimental task, as well as their language and music background (see questionnaire in the Supplementary Materials).

Equipment

Participants were fitted with a custom-built pair of headphones with extra passive sound attenuation (Franken et al., 2019) and an attached DPA 4088-B directional microphone. The microphone sent the recorded signal via a Behringer Xenyx 802 mixing panel to an Eventide Eclipse v4 multi-effects processor, which took care of the actual pitch-shifting. The output was sent, via the mixing panel (where it was mixed with Brownian noise) and an Aphex Headpod 4 amplifier to the custom-built headphones. The volume gain from microphone input to headphone input was kept constant across participants at about a +10dB increase (i.e., the auditory feedback was 10dB above the signal picked up by the microphone). The Brownian noise was on throughout the entire experiment, and its intensity was kept constant at 85dBA. The pitch-shifting algorithm in the Eventide Eclipse v4 is controlled through MIDI messages from a laptop running PsychoPy3 (Peirce et al., 2019). The laptop also records both vocal output and auditory feedback via an external sound card (MOTU MicroBook IIc) with Audacity® recording and editing software (v. 2.3.3).

Analysis

Comparing the results of Experiments 1 and 2 allowed us to investigate the effect of the exposure phase on pitch awareness. In a second stage, the results of Experiments 3 and 4 were analyzed in order to investigate whether any effects observed in Experiments 1 and 2 are due to mere auditory exposure (which should be the same as in Experiments 3 and 4, respectively), or whether they require speech to be produced.

Logistic regression was used to model each participant's responses to the question whether feedback was increased or decreased as a function of the pitch shift magnitude, in the pre-test and the post-test separately. The 50% cut-off of the fitting logistic function was taken as the participant's estimate of their own pitch production, corresponding to the pitch shift value for which the participant would label the

feedback as higher than their own pitch 50% of the time. As exposure to pitch-shifted feedback over the course of the post-test may dissipate any effect of the exposure block, we subdivided the pre- and post-test in subblocks of 30 trials each, to study the temporal stability of the effect (the first 10 trials of the pre-test were considered practice and were not analyzed). For each 30-trial subblock, the 50% cut-off of the logistic fit was calculated for each participant (termed 'pitch awareness' hereafter). For statistical inference, the participants' pitch awareness estimates in the pre- and post-test blocks were entered in a linear mixed effects model, where they were modeled as a function of subblock (each of the 30-trial subblocks) and experiment, with random intercepts across participants.

As the size of these subblocks (i.e., 30 trials) is arbitrary, a separate analysis sought to model the temporal development of responses over time, using a generalized additive model (GAM) with a logit link function. This model type allows to model changes over time without the assumption of linearity (Baayen et al., 2017; Montero-Melis & Jaeger, 2019; Winter & Wieling, 2016), while taking into account random variability by letting intercept, slope, as well as the shape of the relationship (i.e., 'wigglyness' of the curve) vary across participants. However, GAMs do penalize non-linearity in the relationship between predictors and the dependent variable, so that linear relationships are preferred. In the current study, a logistic GAM was used to model the development of participants' responses across trials in the post-test as a function of Pitch Shift and Experiment. This analysis was carried out using the R packages 'mgcv' (Wood, 2017) and 'itsadug' (van Rij et al., 2017).

In addition, the pitch of participants' vowel production was analyzed with the autocorrelation method implemented in Praat (Boersma & Weenink, 2017). In order to remove pitch estimation errors, pitch values that exceeded the threshold of 3 standard deviations above or below the average (per participant) were removed from the analysis. For every vowel utterance, the pitch was estimated at the

vowel center and expressed in cents using the following formula, where pitch$_{Hertz}$ refers to the current utterance's pitch estimated in Hertz.

$$Pitch_{cents} = 1200 * \log_2 \left(\frac{pitch_{Hertz}}{200} \right)$$

Subsequently, the average pitch in each participants' pre-test was subtracted from each trial's pitch estimate, in order to express pitch as a change from the pre-test average, which served as a participant-specific baseline. The resulting pitch estimates for the exposure block were entered in a linear mixed effects model with the fixed effects factor experiment (Experiment 1 vs. Experiment 2). The random effects included random intercepts across participants and across vowels, and by-vowel random slopes for experiment.

Results

Pitch Awareness in Experiments 1 and 2

As the main hypothesis suggests that pitch awareness would be modified by exposure to altered auditory feedback in the exposure block, exposure to altered auditory feedback over the course of the post-test could lead the expected effect to dissipate. Therefore, a first analysis compared participants' responses in the first 30-trial subblock of the post-test, compared to the last 30 trials of the pre-test (Figure 2). It can be observed that in Experiment 1, the logistic curve shifts to the right from pre-test to post-test, while it shifts to the left in Experiment 2. In other words, after prolonged exposure to high-pitched auditory feedback, participants were less likely, in the short term, to label auditory feedback as higher than their actual pitch, but they were more likely to do so after exposure to very low-pitched auditory feedback.
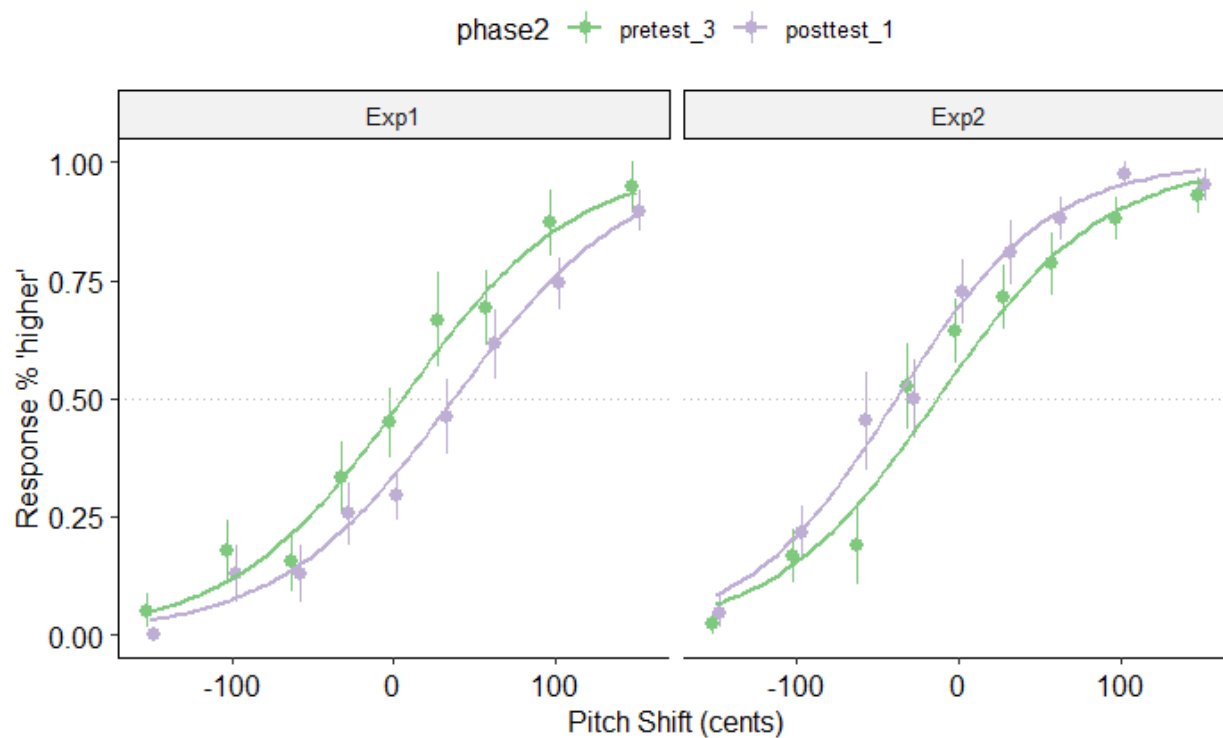
Figure 2. Proportion of 'higher' responses as a function of Pitch Shift in the last 30 trials of the pre-test (green) and the first 30 trials of the post-test (purple). Pitch shift magnitudes include 0, 30, 60, 100 and 150 cents. The left panel illustrates the results for Experiment 1, the right panel for Experiment 2.

Comparing participants' responses across the entire pre- and post-test blocks confirms that this effect

indeed dissipates relatively quickly into the post-test block: there is no difference in pitch awareness

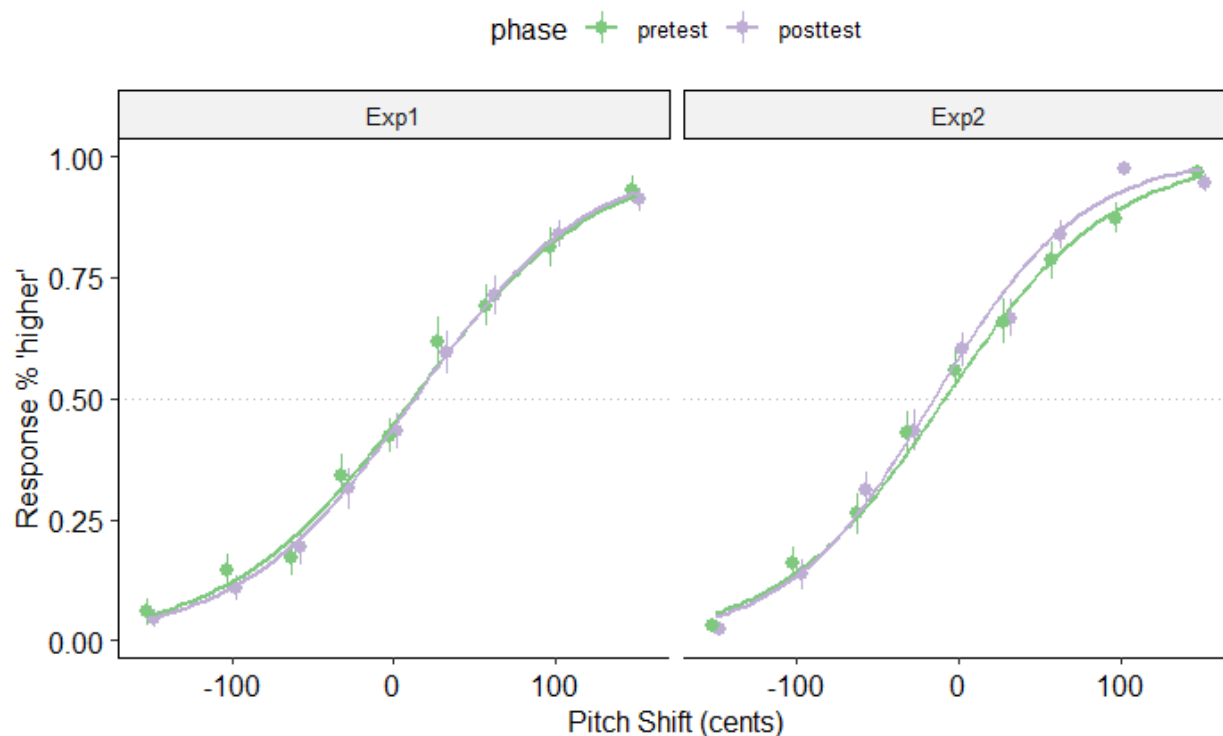when analyzing the entire pre-test and post-test as one block each (Figure 3).

Figure 3. Proportion of 'higher' responses as a function of Pitch Shift in the pre-test (green) and post-test (purple) blocks. The

left panel illustrates the results for Experiment 1, the right panel for Experiment 2.

In order to get a better understanding of the temporal stability of the effect, estimates of participants'

pitch awareness, as measured by the point where the logistic curve crosses the 50% line were calculated

for each 30-trial subblock, illustrated in Figure 4. The linear mixed effects analysis demonstrated a

significant interaction between subblock and experiment ($F(6, 153.08) = 4.29$, $p < .001$), driven by a

significant difference in pitch awareness between Experiment 1 and Experiment 2 of 79.4 cents for the

first subblock of the post-test (est. = 79.41, $\chi^2(1) = 27.74$, $p < .001$, Holm-corrected), while the difference

between experiments was not significant for any other subblock (Table 2). This suggests that

immediately after exposure, participants' pitch awareness was shifted in the direction of the exposure

block's pitch shift, but this effect did not last beyond the initial 30 trials after exposure.
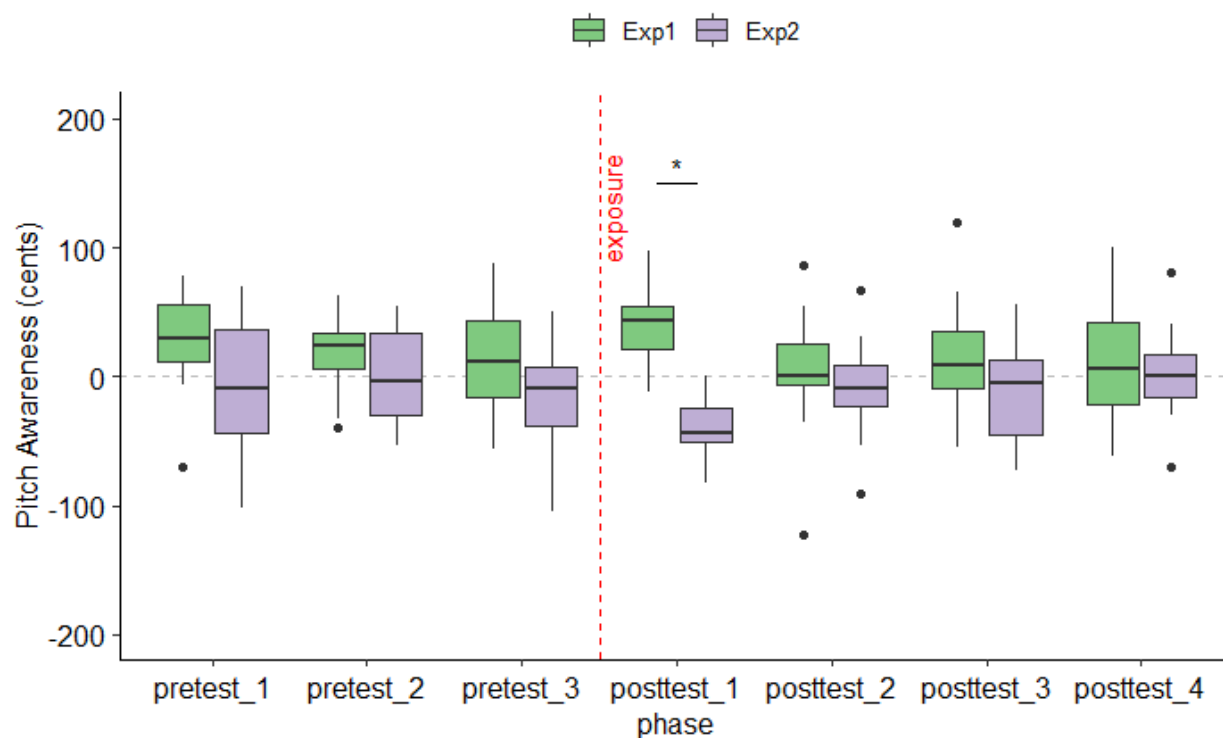

Figure 4. Pitch Awareness in Experiments 1 and 2, expressed as a function of 30-trial subblock and Experiment. The red dashed

vertical line indicates the point where participants were exposed to either the high-pitched (Experiment 1) or the low-pitched

(Experiment 2) feedback voice. The asterisk indicates a significant difference in pitch awareness between Experiments.

Table 2. Difference between pitch awareness of Experiment 1 and Experiment 2 per 30-trial subblock. P-values are adjusted for family-wise error rate using Holm's method.

Subblock	Contrast	Pitch Awareness difference	$\chi^2(1)$	p
pretest 1	Exp 1 − Exp 2	35.00	5.26	.13
pretest 2	Exp 1 − Exp 2	15.85	1.08	.90
pretest 3	Exp 1 − Exp 2	31.62	4.40	.18
posttest 1	**Exp 1 − Exp 2**	**79.41**	**27.74**	**< .001***
posttest 2	Exp 1 − Exp 2	12.52	0.69	.90
posttest 3	Exp 1 − Exp 2	27.78	3.39	.26
posttest 4	Exp 1 − Exp 2	11.33	0.55	.90

As 30 trials is an arbitrary size for the division of the experiment into subblocks, in a further analysis we modeled the responses with a generalized additive mixed model (GAM). This allows us to model how response probability (or its logit-transformed equivalent) varies across trials. The results confirmed that the change in pitch awareness dissipates after about 30 trials (see supplementary materials for details).

Pitch Awareness in Experiments 3 and 4

The main purpose of Experiments 3 and 4 was to further investigate whether production is necessary for the exposure effect in Experiments 1 and 2. In Experiments 3 and 4, participants were only exposed to high- and low-pitched vowels by passive listening, instead of speech production with altered auditory feedback. Figure 5 shows Pitch Awareness as a function of 30-trial subblocks and Experiment. For the sake of comparison, the data for Experiments 1 and 2 are repeated (see Figure 4). In contrast to Experiments 1 and 2, a linear mixed effects model on the pitch awareness estimates in Experiments 3

and 4 revealed no significant interaction between subblock and experiment (F(6, 144.16) = 1.52, p =

.18).While the pitch awareness values for Experiments 1 and 2 differed in the first posttest subblock, no

such difference was detected between Experiments 3 and 4 (est. = 27.71, χ^2(1) = 2.10, p > .99). This

suggests that the passive exposure in these experiments did not lead to a change in pitch awareness.
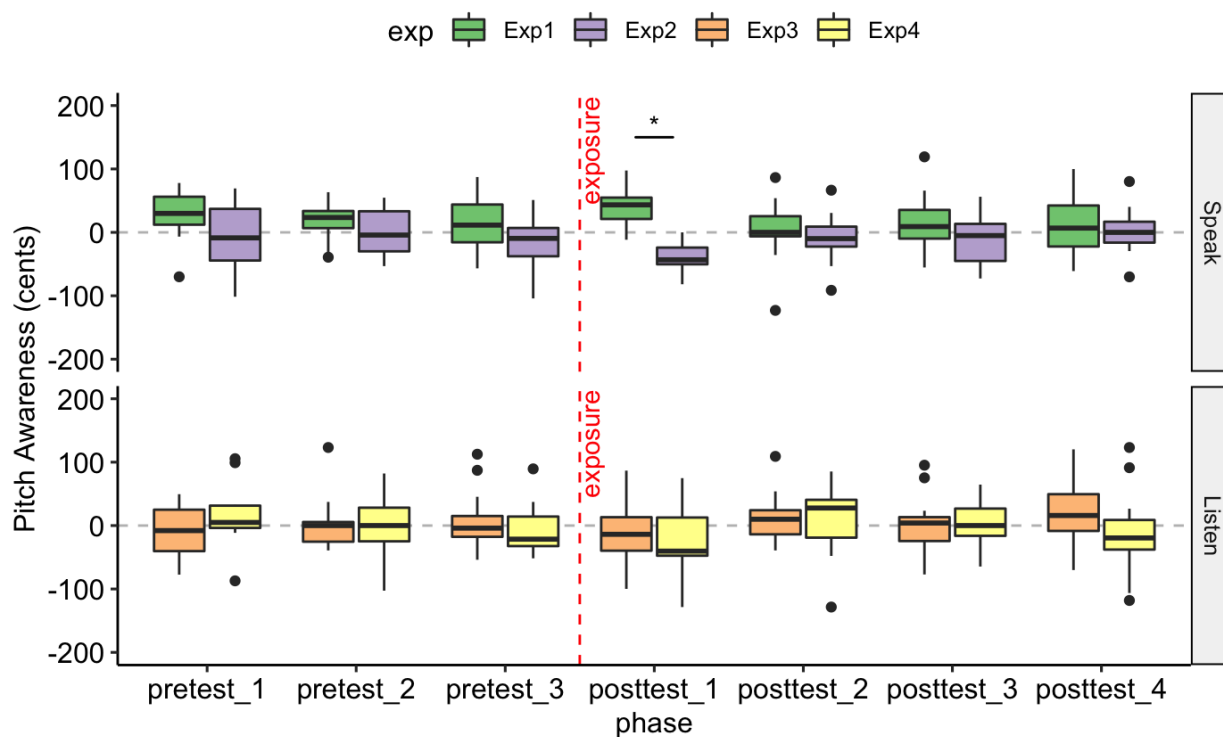

Figure 5. Pitch Awareness expressed as a function of 30-trial subblock and Experiment. The top panel shows the results for

experiments with a speaking task in the exposure block (Experiments 1 and 2), and the bottom panel shows the results for

experiments with a listening task in the exposure block (Experiments 3 and 4). The data for Experiments 1 and 2 is repeated

from Figure 4. The red dashed vertical line indicates the point where participants were exposed to either the high-pitched

(Experiment 1, 3) or the low-pitched (Experiment 2, 4) auditory stimuli.

As we had specific hypotheses about the comparison between Experiments 1 and 2, on the one hand,

and Experiments 3 and 4, on the other hand, a linear mixed effects model was run on the pitch

awareness estimates from the last pretest subblock and the first posttest subblock of all four

experiments, as a function of Subblock and Experiment. In line with the previous analyses, there was a

significant interaction between subblock and experiment (F(3, 51) = 4.03, p = .012). Crucially, we further

investigated as a planned contrast whether the difference between Experiments 1 and 2 was different

from the difference between Experiments 3 and 4. As shown in Table 3, this was indeed the case in the

first subblock of the post-test, suggesting that the exposure phase had a larger effect on pitch

awareness during production than during passive listening.

Table 3. Planned contrasts for across-experiment comparisons of pitch awareness. These were based on the linear mixed

effects model PitchAwareness ~ Experiment*Phase + (1 | Participant). All p-values were Holm-corrected for a family of 2 tests.

Contrast	Subblock	Est.	$\chi^2(1)$	p (Holm-corrected)
(Exp1 − Exp2) − (Exp3 − Exp4)	pretest3	18.51	0.63	.43
	posttest1	56.93	5.99	.029*

Pitch alignment

Participants' pitch productions across Experiments 1 and 2 are shown in Figure 6. While participants'

pitch stayed relatively steady overall, the pitch in the exposure block in Experiment 1 tended to increase,

while in Experiment 2 it tended to decrease. The results of a linear mixed effects model showed that

there was a significant interaction between Experiment and Experimental block ($F(2, 26.01) = 3.61$, p =

.041). Specific contrasts investigated whether the difference in pitch between Experiments varied across

phases (see Table 4). Although not strictly significant, the pitch difference between experiments tended

to be larger during exposure compared to pretest. In addition, the by-experiment pitch difference was

significantly smaller during post-test compared to exposure. Together, these results suggest that

speakers aligned their pitch with the constant pitch-shifted feedback during the exposure block.

However, the magnitude of pitch alignment, as quantified by the participant-specific difference in pitch between exposure and pre-test, was not correlated with the modulation of pitch awareness as quantified by the participant-specific change in pitch awareness in the first 30 trials of the post-test ($r(26) = .184$, $p = .350$). This suggests that the pitch alignment observed in pitch production is not associated with the modulation of pitch awareness.
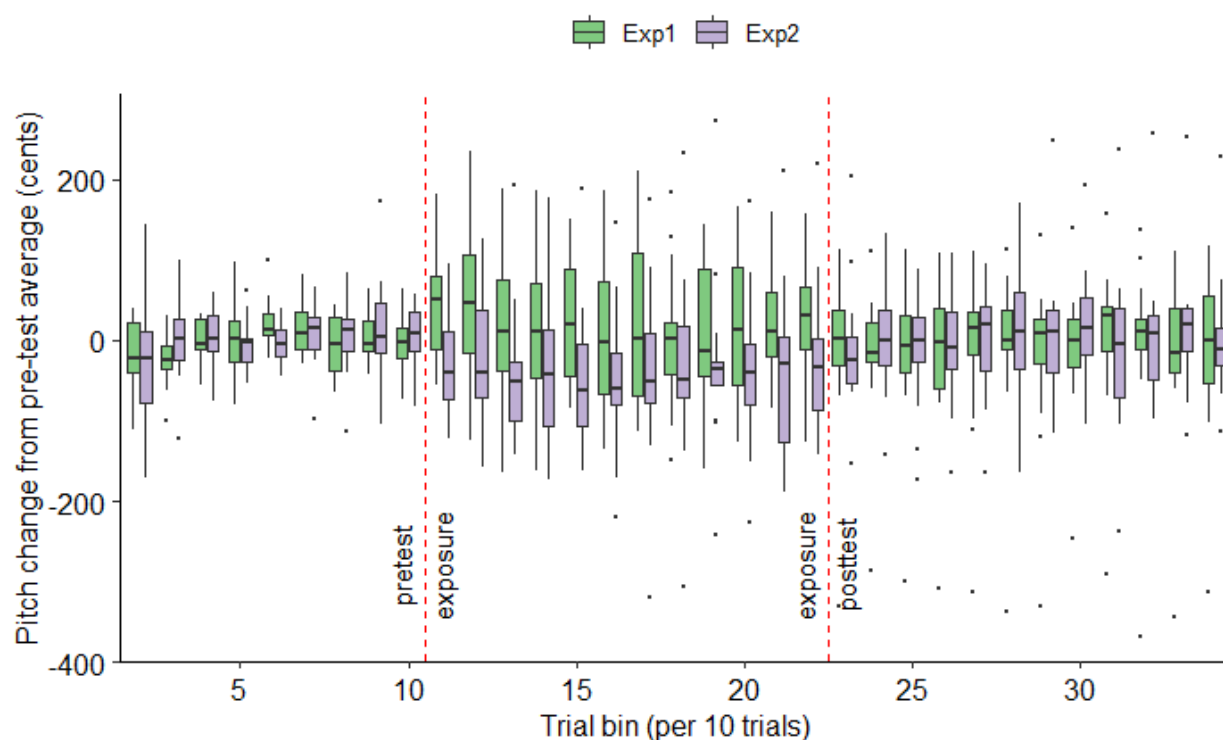

Figure 6. Average produced pitch per 10-trial bin. Colors indicate experiment, and the red dashed vertical lines indicate separations between pre-test and exposure, and between exposure and post-test.

Table 4. Specific contrasts on average pitch across Experimental Phases for each Experiment. P-values are Holm-corrected for multiple comparisons.

Experiment	Block contrast	Estimate	χ^2 (1)	p

| Exp 1 – Exp 2 | exposure – pretest | 55.72 | 3.20 | .074 |
| Exp 1 – Exp 2 | posttest – exposure | -69.82 | 6.51 | .021* |

Verbal reports

In order to explore how participants subjectively experienced the experimental task, a questionnaire was filled out by the participants after the experiment (see Supplementary Materials). The results of these exploratory analyses are visualized in Figure 7. Participants were asked to rate whether they felt they knew their actual pitch during the experiment on a scale from 1 to 10, where a score of 1 indicated "I had no idea about my actual pitch", while 10 indicated "I knew my actual pitch exactly". Figure 7A shows a histogram of participants' ratings as a function of the experimental task in the exposure block (collapsing Experiments 1 and 2 as 'Speaking' and Experiments 3 and 4 as 'Listening'). Participants' ratings in Experiments 1 and 2 seem to be biased towards the low end of the scale. This suggests that in these experiments, participants did not feel they knew their actual pitch well. Furthermore, participants were asked whether they thought the task became easier or more difficult over the course of the experiment. Figure 7B shows a histogram of their responses, where most people in Experiments 1 and 2 felt that the knowledge of their pitch decreased over the course of time, while most participants in Experiments 3 and 4 felt no change. This suggests that the exposure to constant pitch-shifted feedback during speech production (i.e., in the exposure block of Experiments 1 and 2) decreased participants' confidence about their produced pitch.
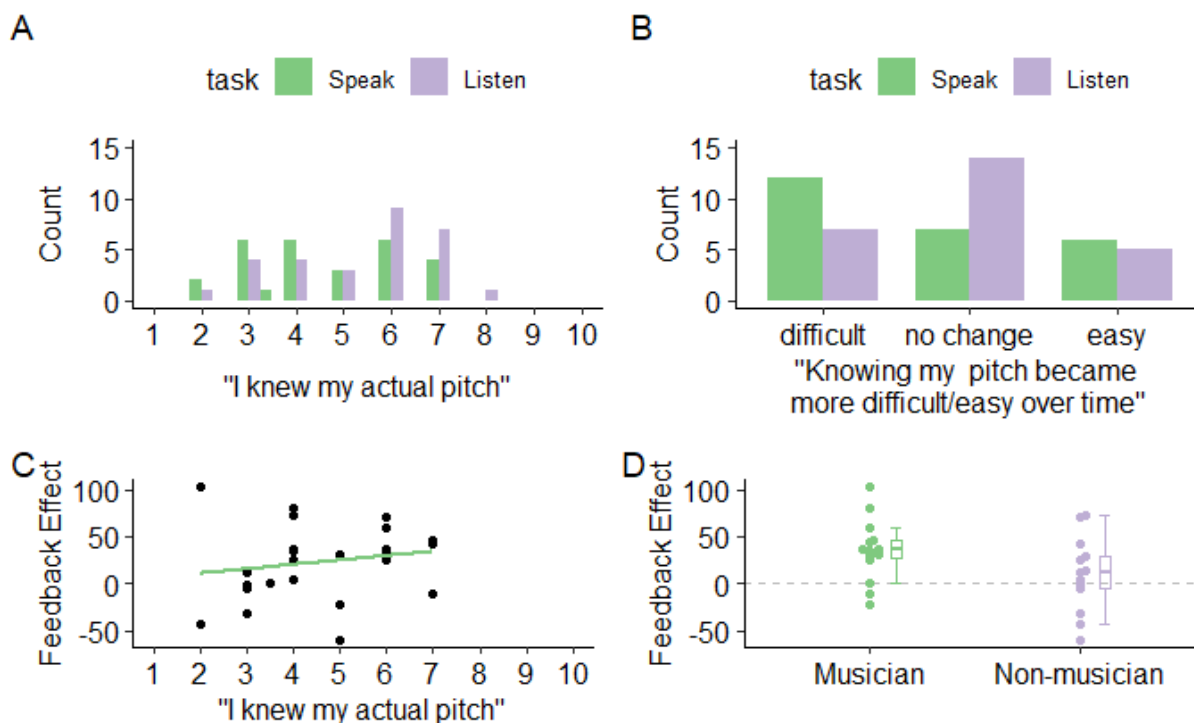
Figure 7. Results of the questionnaire. A. Histogram of rating results on the statement "I knew my actual Pitch" on a 10-point scale, as a function of task in the exposure block. 'Speak' collapses the results from Experiments 1 and 2, while 'Listen' collapses the results from Experiments 3 and 4. The lower and the higher end of the rating scale were labeled respectively as "I have no idea about my actual pitch" and "I know my actual pitch exactly". B. Histogram of participants' responses to the question "Did the extent to which you knew your own pitch change over the course of time?". C. Scatter plot of the effect of auditory feedback on pitch awareness as a function of the rating results (Experiments 1 and 2). The green line indicates the linear fit. D. The effect of feedback on pitch awareness as a function of whether participants had at least some experience with singing or playing a musical instrument (Experiments 1 and 2).

Furthermore, in order to explore the individual variability in the effect of feedback on pitch awareness found in Experiments 1 and 2, the effect of exposure to altered feedback was quantified for each participant as the difference between pitch awareness in the last 30 trials of the pretest and the first 30 trials of the post-test. The sign of the resulting values for participants in Experiment 2 was changed, so

that positive values indicated a change in feedback awareness in the predicted direction, independent of the experiment. These values are compared with some of the results from the questionnaire, shown in Figure 7(C & D). A comparison between the ratings (as shown in Figure 7A) and the feedback effect showed a slight positive trend where participants who felt they knew their actual pitch better than others showed a stronger feedback effect. A comparison between participants who had at least some experience with playing a musical instrument or with singing with those who had no such experience, suggested a trend that musicians showed a stronger feedback effect. However, note that neither of these associations were significant (rating: Spearman's rho = 0.26, p = .18; musicians: $t(25.29)$ = 1.69, p = .10), and given the exploratory nature of these analyses, they should be taken with a grain of salt. Other information that was gathered from the questionnaire, such as the number of languages participants reported to speak, showed no descriptive trends to be associated with the feedback effect.

Discussion

The current study investigated whether internal representations of one's pitch, or pitch awareness, rely on auditory feedback. Four experiments examined whether exposure to high- or low-pitched auditory input led to changes in pitch awareness, and whether this is driven by speech production. If pitch awareness relies on auditory feedback, exposure to pitch-shifted auditory feedback should lead to a drift in pitch awareness.

The results of Experiments 1 and 2 suggest that participants' estimates of their own pitch productions were affected by exposure to pitch-shifted feedback. After exposure to constant high-pitched auditory feedback, participants were less likely to label pitch-shifted feedback as 'higher than their own pitch', which suggests that the constant pitch shift induced them to experience their own productions to be higher in pitch. In contrast, exposure to constant low-pitched auditory feedback made participants more likely to label subsequent feedback as 'higher than their own pitch', suggesting that they judged their

own production to be lower in pitch. These results suggest that speakers do not have a fixed pre-articulatory representation of their vocal pitch productions, but instead, such representations are affected by auditory feedback. The lack of a similar effect in Experiments 3 and 4 show that the change in pitch awareness was driven by speech production with pitch-shifted feedback, rather than by mere perceptual exposure to high- or low-pitched auditory stimuli. In addition, while tentative only, analyses of the verbal reports suggested that most participants in Experiments 1 and 2 felt that they did not know their actual pitch very well, and that this became more difficult over time. In other words, exposure to pitch-shifted feedback in Experiments 1 and 2 led to a decreased confidence in one's pitch awareness. These results should be interpreted with caution however, given their exploratory nature and the lack of statistical significance.

The wide-spread comparator model of the sense of agency suggests that self-generated auditory input is distinguished from auditory input generated by others by comparing an internal representation of one's speech with the observed auditory input. A match would label the auditory stimulation as self-caused, while a mismatch characterizes it as externally generated. The current results, however, are problematic for this account, as they suggest that the internal pitch representation is affected by exposure to pitch-shifted feedback. In response to altered auditory feedback, speakers seemed to alter their internal model. This way, the internal representation cannot serve as a clear benchmark against which auditory input is compared to identify input as self- or other-caused. An alternative account holds that distinguishing between self and others is based not only on a comparison between predicted and observed sensory feedback, but on the integration of multiple sensory cues (Lind et al., 2014a). If other cues in addition to the mismatching pitch in the auditory input still suggest that the auditory input is self-caused, the speaker may want to adjust their internal speech representation, in order to resolve the mismatch with auditory feedback. Relevant other cues may for example include the temporal synchronicity of auditory input and speech production. In addition, participants may vary in their

reliance on internal vs. external cues to determine the source of sensory signals. For example, Synofzik et al. (2010) argued that delusions of control may be based on imprecise internal predictions about the sensory consequences of one's actions. As such imprecision may prompt participants to rely more strongly on external cues such as auditory input, it is an outstanding question to what extent the presence of altered auditory feedback may have prompted participants' reliance on it.

The current study also suggested that participants tended to align pitch with the manipulated feedback during the exposure block. While many studies that use pitch-shifted feedback find that participants compensate for altered feedback, these studies typically use much smaller pitch manipulations (Burnett et al., 1998; Larson et al., 2001). In contrast, a pitch alignment as in the current study was found in several previous studies employing larger pitch discrepancies (Franken et al., 2021; Tajadura-Jiménez et al., 2017; Zheng et al., 2011). It is at present unclear whether this alignment is driven by the sense of agency over the manipulated feedback, although several authors have argued this may be the case (Tajadura-Jiménez et al., 2017; Zheng et al., 2011). The vast literature on speech adaptation shows that with prolonged exposure to small (often unnoticed) feedback manipulations, speakers tend to alter their speech in the opposite direction, effectively compensating for articulatory (Houde & Jordan, 1998; Purcell & Munhall, 2006) or vocal pitch (Jones & Munhall, 2000) manipulations. Studies that made use of brief, unexpected, pitch shifts in the feedback show that with increasing pitch shift magnitude, the probability of a compensatory responses decreases, and the probability of a following response increases (Burnett et al., 1998; Scheerer et al., 2013). Recent studies have suggested that this reduction in compensatory, and increase in following, responses is indicative of a loss of sense of agency for larger shifts (Korzyukov et al., 2017; Subramaniam et al., 2018). However, it is important to note that these studies make use of unpredictable, brief pitch changes, while the current exposure phase uses a constant ±500 cents pitch manipulation. A recent study of ours combined both brief and constant pitch shifts, showing that a constant +500 cent pitch manipulation does not greatly reduce the sense of

agency, as measured by compensatory responses to brief unexpected pitch shifts (Franken et al., 2021). This study showed that speakers aligned their pitch with the constant 500 cents pitch shift, while compensating for brief unexpected smaller pitch shifts. Finally, the lack of a by-subject correlation between this pitch alignment and the change in pitch awareness after exposure, suggests that these are separate processes. So, although prolonged exposure to pitch-shifted feedback led both to a change in pitch awareness and to pitch alignment, there is no evidence that the two processes are associated. This is also supported by the different timelines, where pitch alignment is observed during exposure, but disappears immediately in the post-test block, while the change in pitch awareness can be observed during at least the first 30 trials of the post-test.

What drives participants to estimate their own pitch as lower or higher after the exposure block? The results of Experiments 3 and 4 show that mere perceptual exposure to high- or low-pitched auditory stimuli does not modulate participants' pitch awareness during subsequent speech production. This is in line with the view we have advocated above that the representation speakers have of their own vocal production is influenced by auditory feedback. We suggest that one of the factors why vocal production during exposure is crucial for a modulation of pitch awareness, may be the associated sense of agency. Given that participants produced vowels during exposure in Experiments 1 and 2, we can assume that they had a sense of agency over the auditory signal, while they did not experience agency over the auditory stimuli that they passively listened to during exposure in Experiments 3 and 4.

To some extent, the current study's results are consistent with previous indications that exposure to altered auditory feedback leads to perceptual changes (Lametti et al., 2014). These authors showed that formant perturbations in auditory feedback led to motor adaptation as well as to changes in speakers' perceptual phoneme boundary as measured by an auditory categorization test. In contrast, the current study shows a change in pitch awareness, measured during a speech production task, without evidence

of motor adaptation. While both results suggest that altered auditory feedback leads to changes in the relevant auditory-motor mapping, it is unclear at this point to what extent these are related. There are important differences between these two studies, including formant vs. pitch manipulations, perturbation magnitude, the presence of motor adaptation, and the perceptual task (Lametti et al. used a perceptual categorization task, whereas we asked participants to report on their perception of sensory feedback during a speech production task). In addition, it has been suggested that history of exposure to formant-shifted feedback affects subsequent reliance on auditory feedback (Niziolek & Parrell, 2021). In line with this idea, we find evidence that the history of exposure to auditory feedback affects speakers' internal representations. It is important to note that both Lametti et al. (2014) and Niziolek & Parrell (2021) used formant-shifted feedback. Some previous studies have suggested that there are important differences between vocal pitch control and articulatory motor control such as responses to formant-shifted auditory feedback (Lester-Smith et al., 2020; Max et al., 2003), suggesting we should be cautious in interpreting any parallels or seemingly contradicting findings.

Overall, the current study shows that pitch awareness is flexible, as it can be modified by prolonged exposure to altered auditory feedback. This result challenges the comparator model of agency, which proposes that the sense of agency is generated by a comparison between observed sensory feedback, and an internal representation of the performed action. The flexibility of pitch awareness in the context of altered sensory feedback in the current study suggests that these internal representations cannot act as the sole benchmark for the generation of the sense of agency. While this comparison process may still play a role in generating a sense of agency, the results suggest that other contextual factors should be taken into account, as suggested by inferential models. In addition, future work examining responses to pitch-shifted auditory feedback should consider that internal representations may be affected by the history of altered feedback, and thus possibly in turn could affect subsequent responses.

Acknowledgements

Declarations

Competing interests: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval: All experimental procedures were approved by the ethical committee of the Faculty of Psychology and Educational Sciences of Ghent University, and were designed and carried out in accordance with the Declaration of Helsinki.

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Availability of Data and Code: The datasets generated during and/or analysed during the current study are available on the Open Science Framework, DOI 10.17605/OSF.IO/3KSQ2.

References

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

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

Blakemore, S.-J., 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

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2021). Speaking with an alien voice:

Flexible sense of agency during vocal production. *Journal of Experimental Psychology: Human

Perception and Performance*, *47*(4), 479–494. https://doi.org/10.1037/xhp0000799

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

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

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to

make a voluntary response reveals the presence of two components to the audio-vocal reflex.

Experimental Brain Research, *130*(2), 133–141. https://doi.org/10.1007/s002219900237

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

1213–1216.

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

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

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the

Human Speech Motor System Drives Changes in Speech Perception. *Journal of Neuroscience*,

34(31), 10339–10346. https://doi.org/10.1523/JNEUROSCI.0108-14.2014

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

Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The

Relation of Articulatory and Vocal Auditory–Motor Control in Typical Speakers. *Journal of

Speech, Language, and Hearing Research*, *63*(11), 3628–3642.

https://doi.org/10.1044/2020_JSLHR-20-00192

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014a). Auditory feedback of one's own

voice is used for high-level semantic monitoring: The "self-comprehension" hypothesis. *Frontiers

in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00166

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014b). 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

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

Max, L., Wallace, M. E., & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. *Proceedings of the 15th Congress of Phonetic Sciences*.

Montero-Melis, G., & Jaeger, T. F. (2019). Changing expectations mediate adaptation in L2 production. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/S1366728919000506

Niziolek, C. A., & Parrell, B. (2021). Responses to Auditory Feedback Manipulations in Speech May Be Affected by Previous Exposure to Auditory Errors. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2169–2181. https://doi.org/10.1044/2020_JSLHR-20-00263

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Scheerer, N. E., Liu, H., & Jones, J. A. (2013). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *The European Journal of Neuroscience*, *38*(8), 3189–3200. https://doi.org/10.1111/ejn.12301

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds

    following speech motor learning. *J Acoust Soc Am*, *125*(2), 1103–1113.

    https://doi.org/10.1121/1.3058638

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., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in

    schizophrenia are based on imprecise predictions about the sensory consequences of one's

    actions. *Brain*, *133*(1), 262–271. https://doi.org/10.1093/brain/awp291

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

van Rij, J., Wieling, M., Baayen, H., & van Rijn, H. (2017). *itsadug: Interpreting Time Series and*

    Autocorrelated Data Using GAMMs (R package version 2.3).

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve

    Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, *1*(1), 7–18.

    https://doi.org/10.1093/jole/lzv003

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

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (Second Ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

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