Does passive sound attenuation affect responses to pitch-shifted auditory feedback?

Matthias K. Franken[1], Robert J. Hartsuiker

Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

Petter Johansson, Lars Hall

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

Tijmen Wartenberg

Hearing Technology @ WAVES, Information Technology, Ghent University, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium

Andreas Lind

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

[1] Electronic mail: matthias.franken@ugent.be

Abstract

The role of auditory feedback in vocal production has mainly been investigated by altering auditory feedback (AAF) in real time. In response, speakers compensate by shifting their speech output in the opposite direction. Current theory suggests this is caused by a mismatch between expected and observed feedback. A methodological issue is the difficulty to fully isolate the speaker's hearing so that only AAF is presented to their ears. As a result, participants may be presented with two simultaneous signals. If this is true, an alternative explanation is that responses to AAF depend on the contrast between the manipulated and the non-manipulated feedback. This hypothesis was tested by varying the passive sound attenuation (PSA). Participants vocalized while auditory feedback was unexpectedly pitch-shifted. The feedback was played through three pairs of headphones, with varying amounts of PSA. The participants' responses were not affected by the different levels of PSA. This suggests that across all three headphones, PSA is either good enough to make the manipulated feedback dominant, or that differences in PSA are too small to affect the contribution of non-manipulated feedback. Overall, the results suggest that it is important to realize that non-manipulated auditory feedback could affect responses to AAF.

I. INTRODUCTION

An influential technique for investigating the interplay between speech and auditory feedback is to alter auditory feedback in real time so that speakers hear their productions perturbed in various ways (e.g., in pitch or formants). The dominant view in the field holds that speakers usually compensate for feedback perturbations of pitch and formants because they try to minimize the discrepancy between an internal representation of the sensory speech target and the perceived auditory feedback (Hain et al., 2000; Liu and Larson, 2007). This view, however, ignores a methodological issue associated with the altered auditory feedback (AAF) technique: it is very difficult to completely rule out that speakers still perceive their original, unperturbed feedback in addition to the manipulated signal. Thus, it is possible that speakers receive conflicting evidence of what they are producing: their actual, unperturbed, auditory feedback, and the altered auditory feedback provided by the researchers. If so, an alternative explanation for compensation responses is that compensatory responses depend on the conflict between two simultaneous auditory feedback signals. The speaker, in the assumption that the dominant, manipulated, feedback is self-produced, tries to minimize the discrepancy between the manipulated and the original feedback, which leaks through the headphones and is considered as an external reference in this scenario. The current study aims to test this alternative hypothesis.

Speakers receive both somatosensory as well as auditory feedback during speech production. Auditory feedback is composed of both air-conducted and bone-conducted feedback. While it is important to acknowledge the contribution of somatosensory feedback and bone-conducted auditory feedback during speech production, the current study focuses explicitly on air-conducted auditory feedback. The study and manipulation of (air-conducted) auditory feedback through AAF has strongly advanced the field of speech motor control.

Studies using this technique have led to several theoretical frameworks for speech motor control (Guenther, 2016; Houde and Nagarajan, 2011). In experiments that make use of AAF, participants are instructed to speak, while their speech is being recorded with a microphone and played back to them, near-simultaneously, through headphones. The experimenters take control of the auditory feedback by manipulating it in real time, creating a discrepancy between speech intent and the observed auditory signal. The type of manipulations that have been applied include shifting the pitch (Burnett et al., 1998; Elman, 1981), formant values (Houde and Jordan, 1998; Purcell and Munhall, 2006), or fricative noise (Casserly, 2011; Shiller et al., 2009) of the speech signal.

Most of these studies use one of two common paradigms. The first paradigm (‘adaptation’) focuses on how speech production is affected after being exposed to altered auditory feedback that is consistently altered in a specific manner. For example, when the value of the first formant (F1) in the auditory feedback was gradually shifted upwards over the course of an experiment, speakers responded by shifting the F1 in their speech in the opposite way (i.e., downwards), and vice versa (Houde and Jordan, 1998; Jones and Munhall, 2000; Purcell and Munhall, 2006). These studies suggest that over time, speakers adapted to consistently altered auditory feedback by changing their feedforward speech motor commands (Franken et al., 2019). The second paradigm (‘compensation’) is aimed at investigating how speakers respond to brief, unexpected changes in auditory feedback during speech production. The present study makes use of this second AAF paradigm, in order to investigate the effect of passive sound attenuation on immediate responses to unexpected auditory feedback. This also allows us to investigate responses to feedback perturbations of different magnitudes and directions. This is in contrast with the earlier study by Mitsuya & Purcell (2016), where the ‘adaptation’ paradigm was used to investigate adaptation to formant manipulations with either insert earphones or circumaural headphones. While the authors concluded that the headphone type did not affect the adaptation results, it is possible that headphone types will affect immediate responses. This is viable since some recent studies have argued that compensation and adaptation are in fact distinct processes (Franken et al., 2019; Parrell et al., 2017).

In the ‘compensation’ paradigm, speakers usually compensate for the altered feedback by shifting their speech production in the opposite direction (Burnett et al., 1998; Hain et al., 2000). For example, when pitch in the auditory feedback was shifted up, participants responded by lowering their pitch, or vice versa. Interestingly, sometimes speakers may also follow the feedback, by changing their speech in the same direction as the feedback manipulation (Behroozmand et al., 2012; Franken et al., 2018a; Patel et al., 2014). Currently, it is unclear what causes following responses, but multiple factors may play a role. Some authors have suggested that following responses indicate that the feedback manipulation is not considered to be self-generated but treated as an external referent, similar to a singer trying to match the pitch of, for example, an accompanying piano (Hain et al., 2000; Patel et al., 2014). Others have suggested following responses might have to do with the velocity of the pitch shift (Guenther, 2016). A recent study has shown that the current state of the speech system (i.e., ongoing pitch fluctuations) may affect whether a speaker opposes or follows a pitch shift (Franken et al., 2018a). The neural correlates of following responses are poorly understood, but recent studies claim that different neural mechanisms may underlie following and opposing responses (Franken et al., 2018b; Li et al., 2013).

A methodological issue with AAF is that it is very difficult to fully isolate the speaker’s hearing so that only the altered feedback is presented to their ears. Many research groups make use of commercial headphones (see Table 1 for a few examples) and these vary in how much passive sound isolation they offer. As a result, the speaker may be presented with two simultaneous auditory feedback signals: a) What they are actually uttering (the original speech signal), and b) what is relayed through the headphones (the manipulated signal). While two simultaneous auditory signals can be perceived as a single blended signal (Alain, 2007), small discrepancies, for example in pitch, may lead to a perception of two separate signals. For example, perception of two simultaneous vowels is aided by small pitch differences between the two vowels (Darwin, 1997; Darwin et al., 2003).

Table 1. Overview of different headphones used in published perturbation studies.

Headphones	Type	Attenuation	Example studies
AKG boomset (K 270 H/C)	Circumaural	NA	Hain et al. (2000), Exp. Brain Research; Liu et al. (2010b), JASA
Etymotic Research ER	Insert Earphones	>30 dB	Cai et al. (2010), JASA ; Behroozmand et al. (2012), JASA
Sennheiser HD 280 Pro	Circumaural	up to 32dB	Franken et al. (2018a), PBR; Keough et al. (2009), JASA
BeyerDynamic DT 770 Pro	Circumaural	18 dBA	Schuerman et al. (2017), JASA;
Stax SR001-MK2	Insert Earphones	NA	Lametti et al. (2012, 2014), J of Neuroscience

Koss ESP950	Circumaural	NA	Flagmeier et al. (2014), Brain and Language; Behroozmand et al. (2015), NeuroImage

Thus, with low passive sound attenuation, the speaker could receive two conflicting sets of evidence about what they are saying. Most studies increase the volume or add noise to the manipulated signal to make it dominant over the original signal. However, it is very difficult to completely rule out that participants still hear their original speech output. As the presence of the original speech signal is often ignored, it is unclear how its potential interaction with the manipulated signal may have affected the results in many of these studies.

The present study aims at investigating the impact of sound attenuation observed in typical headphones used in AAF experiments. Note that while we acknowledge that sound attenuation has no impact on the contribution of bone-conducted auditory feedback, it will affect the level of air-conducted auditory feedback leaking through the headphones, and thus the overall level of non-manipulated auditory feedback. We first established the passive sound attenuation offered by a number of different headphones (Experiment 1), and then carried out an altered auditory feedback experiment (Experiment 2). This allowed us to evaluate the effect of varying sound attenuation degree of different headphones on the response to unexpectedly altered auditory feedback. Specifically, we aim to answer two questions: (1) will passive sound attenuation affect the magnitude of the compensatory response, and (2) will passive sound attenuation affect the likelihood of opposing responses?

The dominant view in the literature is reflected in models that suggest that compensatory responses to altered feedback arise in order to minimize the discrepancy between the intended speech target and the observed feedback signal (Guenther, 2016; Hain et al., 2000; Houde and Nagarajan, 2011). If both the manipulated and the original feedback signals are present, increased passive sound attenuation would make the manipulated signal more dominant compared to the original feedback and thus make the discrepancy between intended pitch and manipulated pitch more salient. Therefore, based on the dominant theoretical framework, we would expect that increased sound attenuation leads to stronger or more compensatory responses.

Alternatively, compensation could depend on the speaker hearing not only the perturbed feedback, but also the (non-perturbed) normal feedback leaking through the headphones. In other words, instead of an internal pitch target as the referent, the non-manipulated auditory signal is considered the referent, as it is the louder of the two auditory feedback signals. Compensatory behaviour could thus be a consequence of the perceived mismatch between the two conflicting auditory signals that the speaker receives. This hypothesis assumes that speakers consider the manipulated feedback as self-produced, and thus try to minimize the mismatch by bringing this signal closer to the original ("actual") feedback that leaks through the headphones. This would suggest that the intended speech target (or an internal forward model prediction) plays a smaller role than often assumed, in line with views that speech production targets are less well defined than most models hypothesize, as it has been proposed for semantic aspects of language production by inferential models (Lind et al., 2014). Note that we don't claim that speakers should be consciously aware of the presence of two simultaneous auditory signals. Previous studies have shown that responses to pitch-shifted feedback occurs automatically, even when instructed not to (Hain et al., 2000). If this alternative hypothesis is true, increased passive sound attenuation should lead to smaller compensations because this would decrease the saliency of the conflict between the two auditory signals. With increased sound attenuation, there would be less sound leaking through the headphones and thus the original, non-manipulated feedback would be less salient, thereby reducing the conflict between two simultaneous feedback signals.

Recent studies have shown that participants sometimes follow and sometimes oppose pitch-shifted feedback (Behroozmand et al., 2012; Franken et al., 2018a). The alternative hypothesis proposed here also provides a more straightforward account for the presence of both opposing and following responses: if two simultaneous signals are perceived, the response direction may depend on which of the two signals is considered by the participant as under their control. To test this hypothesis, we ask whether the proportion of opposing responses might be affected by sound attenuation. While different explanations have been offered to explain following responses, an explanation based on source monitoring of the auditory input as presented here is similar to the account by Hain et al. (2000), who suggest that following might be appropriate when the speaker considers the incoming auditory signal as externally generated, instead of being self-produced (Patel et al., 2014). If this is the case, low passive sound attenuation and thus the presence of two simultaneous auditory signals could lower the probability that the participant will consider the manipulated feedback signal as self-produced, and thus increase the likelihood of a following rather than an opposing response. Thus, this view would predict that increased sound attenuation would lead to a higher proportion of opposing responses.

A recent study investigated a related, but different question (Mitsuya and Purcell, 2016). In order to investigate the role of the occlusion effect, the authors compared insert earphones with circumaural headphones in an adaptation paradigm, and found no effect of headphones on F1 adaptation. In other words, adaptation over time to a consistent manipulation of F1 was not affected by the type of headphones. There is evidence , however, for the hypothesis that longer-term adaptation and immediate compensation to unexpected feedback perturbations may be supported by two different mechanisms (Franken et al., 2019; Parrell et al., 2017). If that hypothesis is correct, the type of headphones could affect these processes in different ways. The current study will focus on real-time compensation responses to unexpected feedback perturbations. In addition, the earlier study compared two headphones in order to examine the role of the occlusion effect. Although it is likely that the headphones used also differed in sound attenuation, the current study will look at the effect of sound attenuation specifically in a pitch-shift compensation paradigm.

II. EXPERIMENT 1: HEADPHONES MEASUREMENTS

In Experiment 1, we investigated the passive sound attenuation of one set of hearing protection ear muffs and seven pairs of headphones. The goal was to have a comparable measure of passive sound attenuation for each pair of headphones, in order to be able to investigate its effect on responses to altered auditory feedback in Experiment 2. Although headphones manufacturers provide sound attenuation measures, it is unclear what method different manufacturers use and thus how these numbers could be compared across headphones. In addition, we measured each headphones' frequency response to make sure differences between the headphones' frequency responses were not a contributing factor to the behavioural differences in Experiment 2.

A. Methods

1. Headphones

Four pairs of commercially available headphones were selected, as well as one pair of hearing protection ear muffs. The headphones were chosen as they were all designed to have high sound attenuation, and reflect the range of headphones commonly used for speech manipulation research. The headphones included three closed-back circumaural headphones, designed to have high passive sound attenuation, as well as the ER-3C insert earphones (Etymotic Research), designed for research. The headphones are listed in Table 2, along with the average attenuation magnitude as specified by the manufacturer. Many of the headphones selected have been used in altered auditory feedback studies (see also Table 1).

Table 2. Attenuation of headphones/ear muffs used in Experiment 1.

Name	Type	Attenuation (according to manufacturer)
Peltor X5A	Hearing protection	37 dB
Beyerdynamic DT 770 Pro	Closed-back circumaural	18 dBA
Sennheiser HD 280 Pro	Closed-back circumaural	up to 32 dB
Vic Firth SIH1	Closed-back circumaural	24 dB
Etymotic ER-3C	Insert earphones	over 30 dB
Custom-built no. 1	Closed-back circumaural	-
Custom-built no. 2	Closed-back circumaural	-
Custom-built no. 3	Closed-back circumaural	-

In addition to the commercially available headphones, we custom-built headphones by placing the loudspeakers (including their plastic casings) of Sennheiser HD 280 Pro headphones into Peltor 3M X5A hearing protection ear muffs[1] (see supplementary materials[2]). Since these custom-built headsets are not standardized, we built three copies of the same design in order to see how they compare to each other. We included the hearing protection ear muffs in our attenuation measurements to check how the construction of the custom-built headphones affected the passive sound attenuation of the ear muffs. The custom-built headphones were created in order to maximize sound attenuation with circumaural headphones. While insert earphones could lead to better sound attenuation still, circumaural headphones avoid an occlusion effect (Mitsuya and Purcell, 2016) and are easier to use.

2. Equipment

For this study we used a Head And Torso Simulator (HATS) Type 4128-C placed in a near-anechoic chamber (only the floor is not anechoic). The HATS is a model of a head and torso designed for in-situ electroacoustic tests. It has models for the human pinnae. However, in the current study the pinnae models were not used as they might interact with some of the circumaural headphones (except for the measurements of the Etymotic Research ER-3C insert earphones, where the pinnae were used). The HATS contains ear simulators with ½’’ microphones, which allows the researcher to record the sound reaching the ears. For the attenuation measurements, acoustic stimuli were played from a single ADAM S1X Active Studio Monitor placed at 1.5m in front of the HATS. At about 2.5cm in front of the mouth of the HATS, a reference microphone (1/2” preprolarized free-field microphone, Bruel & Kjaer type 4189) was placed. Microphones and speakers were connected to a Bruel & Kjaer Input/Output Module (Type 3109).

3. Sound materials

For the measurements of passive sound attenuation, a white noise stimulus was created using Praat (Boersma and Weenink, 2017). In addition, for the frequency response measurements we created stimuli with a male and a female speech-weighted speech spectrum by taking the male-weighted and female-weighted speech-modulated noises from the ICRA project (Dreschler et al., 2001) and randomly shifting the phases in Matlab (Mathworks. Inc., R2016b). All stimuli had a duration of at least 25s.

4. Procedure

In order to measure passive sound attenuation, each pair of headphones was placed on the HATS, while white noise was played at 80 dB SPL through the studio monitor (measured at the reference microphone in front of the HATS mouth). PULSE LabShop (Bruel & Kjaer, v. 15.1.0) was used to control stimulus playback and to record signals from the in-ear microphones as well as from the reference microphone in front of the HATS mouth. Every measurement with the headphones was carried out twice, with repositioning the headphones in between measurements to check for accuracy. The signals were transformed to power spectra (in Pa^2) with 1/3 octave filter bands by averaging over a 20s time window. Before every measurement with headphones, a measurement was carried out without headphones to serve as a baseline measurement. The reference microphone in front of the HATS mouth was used to control the stimulus volume across measurements online. In addition, an offline analysis confirmed that the reference signal was not affected by the presence or absence of headphones on the HATS.

For the measurements of the frequency responses of the headphones, acoustic stimuli were played through each pair of headphones after it was placed on the HATS. Before each measurement, it was made sure that the overall intensity level reaching the in-ear microphones when the headphone was not mounted on the head was 80 dB SPL. Every measurement was carried out twice with headphones repositioning in between. These measurements were repeated with the white noise and the two speech-weighted stimuli.

5. Analysis

The data and analysis scripts are publicly accessible via https://osf.io/vm84u/. All further analyses were done in R (R Core Team, 2018) and focused on frequency bands ranging from 100Hz to 8kHz, which include the frequencies most relevant for speech. The power spectra were expressed in dBA. In order to calculate the attenuation in each frequency band, the intensity in the corresponding frequency band in the baseline measurement without headphones was subtracted from the intensity in the measurements with headphones. This was done for both headphones measurements, after which the results were averaged.

In order to quantitatively compare frequency responses to each other, two metrics were used: spectral flatness, and the average root-mean-square error (Breebaart, 2017). Spectral flatness was quantified as the dB-scaled ratio between the geometric and the arithmetic mean of the power spectrum (Johnston, 1988):

$$Spectral\ Flatness = 10 \cdot \log_{10}\left(\frac{\sqrt[N]{\prod_{n=1}^{N} x(n)}}{\frac{1}{N}\sum_{n=1}^{N} x(n)}\right)$$

Where N is the number of frequency bands, and x(n) the power in frequency band n. The spectral flatness measure has been used to quantify how flat (or noise-like) a spectrum is. It is bounded between -∞ and 0. Given white noise as an input signal, a higher spectral flatness score would thus indicate a frequency response that is closer to the input signal. Only with a perfectly flat spectrum is the geometric mean equal to the arithmetic mean and thus the spectral flatness score 0. Furthermore, frequency responses can be compared to each other by looking at the root-mean-square error (RMSE) between two frequency responses. The RMSE was calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_1(n) - x_2(n))^2}$$

Where $x_i(n)$ indicates the power at frequency band n for headphones i. Both the correlation coefficient and the RMSE value were calculated for each possible pair of headphones, and averaged per headphone. The resulting average values indicate how well a pair of headphones' frequency response compares on average to all the other pairs of headphones. A low value indicates that the frequency response of the headphones is very similar to the other headphones' frequency responses.

B. Results

Figure 1 shows the passive sound attenuation over the frequency range 100-8000Hz for each pair of headphones for both the left and right ears. It is clear from the figure that the passive sound attenuation varies across the frequency spectrum as well as headphones. Note that the Etymotic ER insert earphones seem to be the most attenuating below 300Hz and above 3000Hz, while the hearing protection ear muffs are the most attenuating between 300 and 1600Hz. The different shape of the ER attenuation spectrum compared to the other headphones could be due to the fact that these are the only insert earphones compared to the other (circumaural) headphones, possibly leading to different in-ear resonance frequencies. The fact that we used the HATS' pinna models for the ER measurements but not for the circumaural headphones measurements could be an additional contributing factor. However, for measurements conducted without headphones, the addition of the pinnae models only lead to a slight amplitude increase between 2000 and 5000Hz, suggesting that the pinnae were not a major contributing factor to the spectral differences observed between the ER and the other headphones.

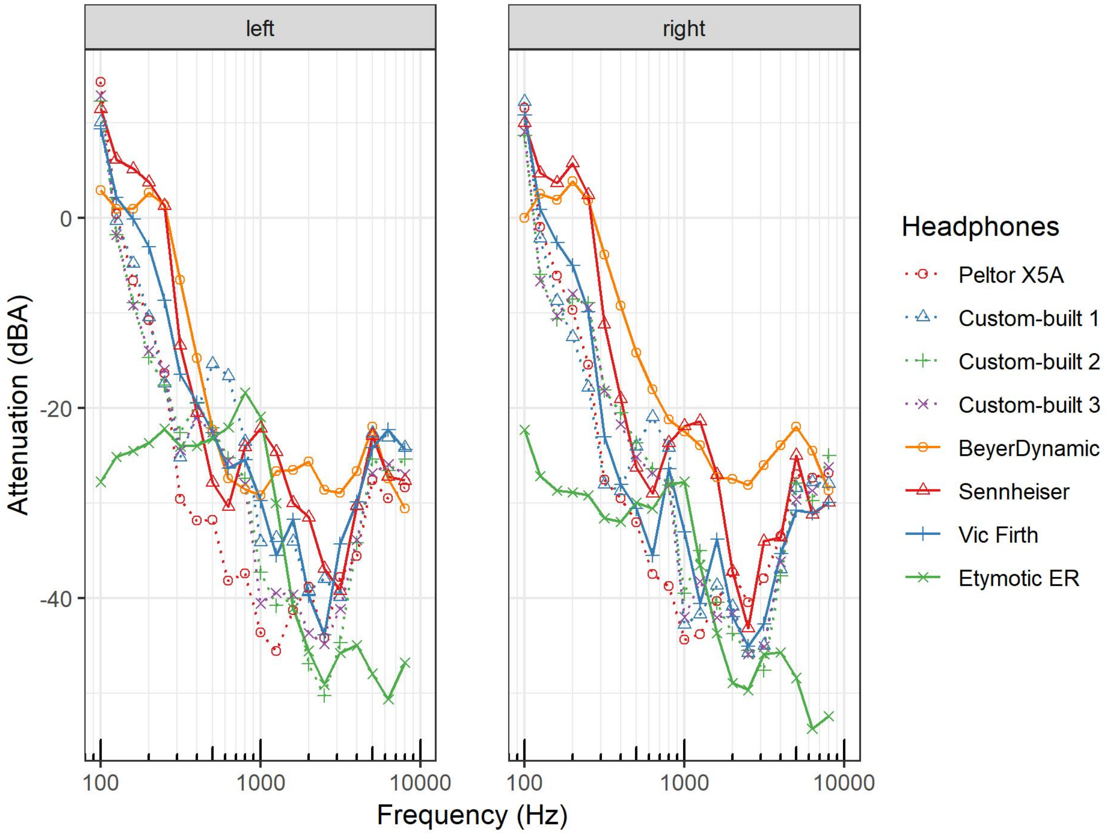

Fig. 1. The measured passive sound attenuation over the frequency spectrum of 100-8000Hz for both left and right ears.

Figure 2 shows the same data, this time averaged across the frequency range, which allows for an overall measure of passive sound attenuation in speech-relevant frequencies. It can be seen from both figures that passive sound attenuation varies across headphones, from the pair of BeyerDynamic and Sennheiser headphones with relatively low attenuation to the most attenuation in the hearing protection ear muffs (Peltor X5A) and the Etymotic ER insert earphones. These values do not precisely correspond to the values provided by the manufacturers as shown in Table 2. A comparison between headphones based on the manufacturer-provided values is difficult, as manufacturers do not disclose how they arrived at these values, and different manufacturers may use different measuring methodologies.

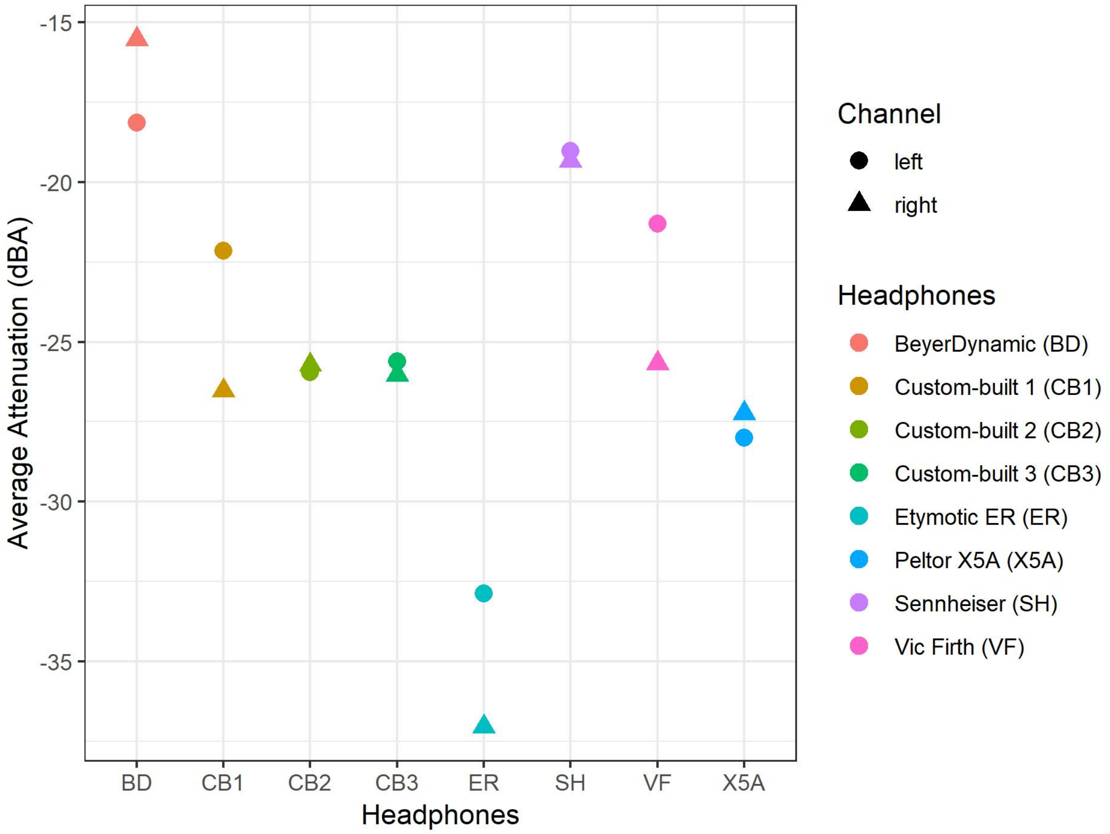

Fig. 2. Average sound attenuation for each pair of headphones, averaged across the 100-8000Hz frequency range.

In order to make sure that any headphones-specific differences in passive sound attenuation are not confounded with headphones-specific frequency response characteristics, the frequency spectrum was quantified for each pair of headphones. First, the spectral flatness of the frequency response to white noise input was quantified for each pair of headphones, shown in Figure 3. Judging from the figure, there is a clear difference in spectral flatness between the Vic Firth headphones and the other headphones. In addition, we see a smaller difference for the left channel between the custom-built pair of headphones nr. 1 and the other custom-built pair of headphones.

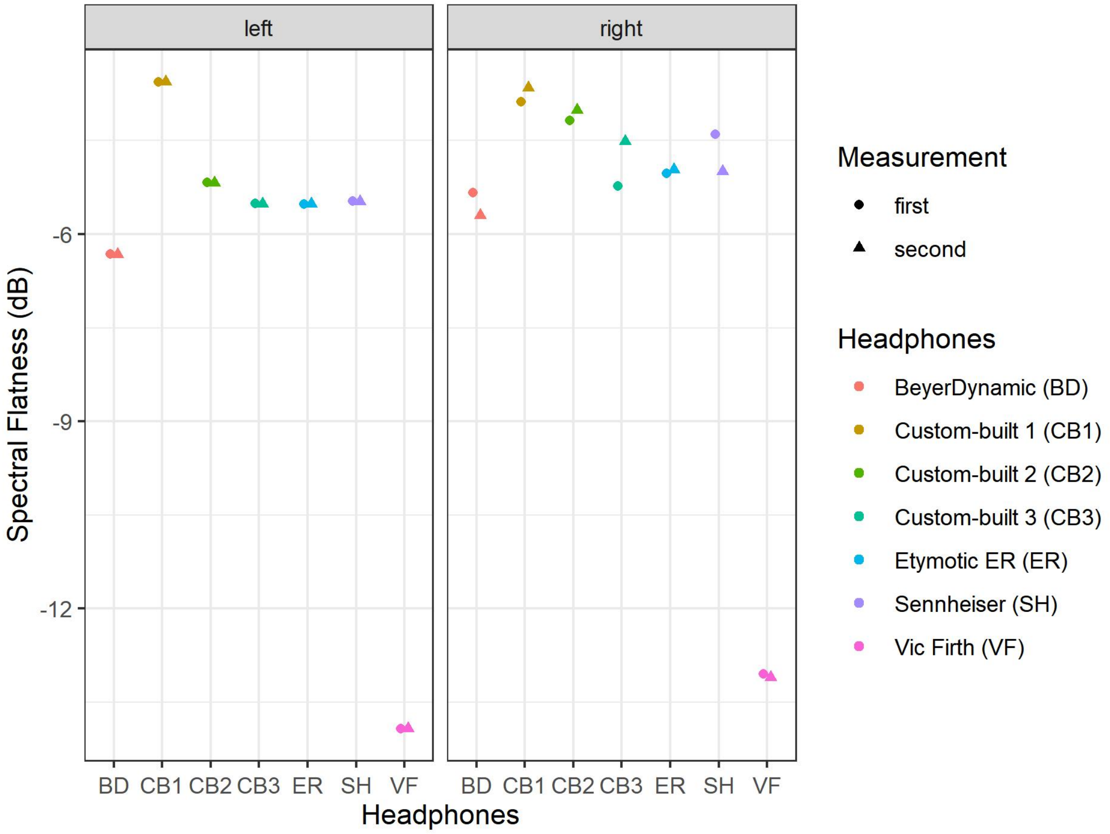

Fig. 3. Spectral flatness of the headphones' frequency response to a white noise (WN) input signal.

A second way to evaluate the differences between headphones' frequency responses is to quantify the average RMSE between a headphones' frequency response and the response of every other pair of headphones. This was done for the frequency response to a white noise input signal, as well as for responses to male (ICRA4) and female (ICRA5) speech-weighted noises, shown in Figure 4. The figure suggests that the Etymotic ER, the custom-built headphones nr. 1, and the Vic Firth headphones show a frequency response which is considerably different from the other headphones.

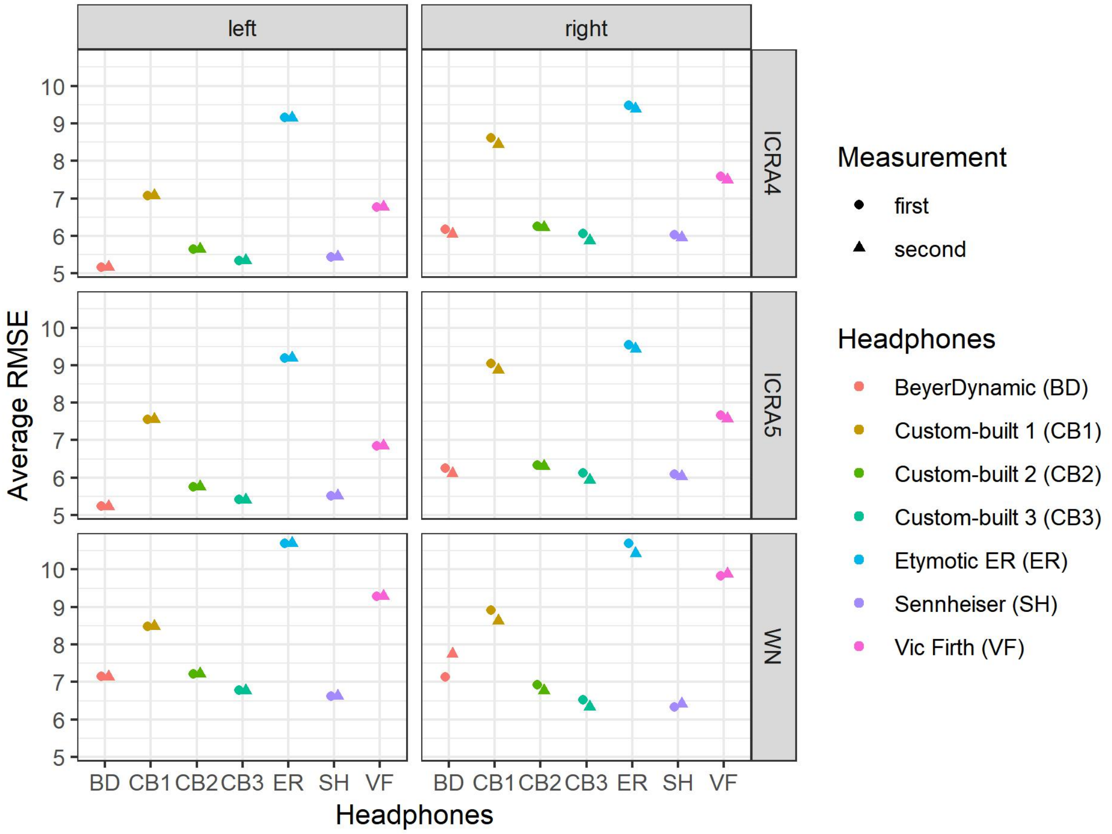

Fig. 4. Average RMSE of every headphones' frequency response.

C. Discussion

Overall, it can be concluded that the pairs of headphones tested in Experiment 1 show variable passive sound attenuation. The BeyerDynamic and Sennheiser headphones show the least sound attenuation, while the Etymotic insert earphones show the highest sound attenuation. The custom-built headphones show medium sound attenuation, approaching the values of hearing protection ear muffs. It should be noted that the high sound attenuation in the Etymotic insert earphones is visible especially for low (<300Hz) and higher (>3000Hz) frequencies. This is interesting in light of the evidence that in speech perception important phonetic cues are conveyed between 100 and about 2000Hz (Epstein et al., 1968; Warren et al., 1995), although higher frequencies also convey speech-relevant information(e.g., for the perception of sibilants). So this result shows that in the ER earphones, attenuation is especially strong in frequencies that are less relevant for many speech sounds.

In order to maximize the range of passive sound attenuation while limiting the number of headphones used, three pairs of headphones spanning the attenuation scale were selected for use in Experiment 2: the BeyerDynamic headphones, a pair of custom-built headphones (Nr. 3), and the Etymotic ER insert earphones. The BeyerDynamic are the least sound-attenuating, the custom-built headphones offer intermediate attenuation, and the ER offer most sound attenuation. This will allow us to interpret differences between headphones in Experiment 2 as a function of passive sound attenuation. Both the BeyerDynamic headphones and the Etymotic insert earphones have been used for altered auditory feedback research in the past (see Table 1). It should be noted that the Etymotic ER insert earphones are somewhat different from the other two, both in type (insert earphones vs. circumaural headphones) as well as in the measured frequency responses (Figure 4). The different frequency response for the ER could affect both the air-conducted auditory feedback, as well as the relative contributions of air-conducted and bone-conducted feedback to the overall auditory feedback, as these differ across the frequency range

(Pörschmann, 2000). This suggests we should take caution interpreting differences between condition with ER earphones and the other two pairs of headphones in Experiment 2 as being solely due to passive sound attenuation.

Finally, Experiment 1 shows that the construction of custom-built headphones by placing Sennheiser headphones speakers into Peltor X5A hearing protection ear muffs was successful, especially for pairs nr. 2 and nr. 3. They showed passive sound attenuation that was not far from the attenuation measured for the Peltor X5A ear muffs, and their frequency response measures were similar to the frequency response of the Sennheiser headphones from which they were constructed.

III. EXPERIMENT 2

In order to investigate whether passive sound attenuation has an effect on speakers' behaviour in a feedback perturbation experiment, three pairs of headphones were selected based on their sound attenuation properties as measured in Experiment 1. Participants took part in a pitch perturbation experiment with three blocks, one for each pair of headphones. If responses to pitch perturbations depend on a comparison between the manipulated feedback and an internal target representation, increased sound attenuation should lead to stronger opposing responses (compared to weaker sound attenuation). If, on the other hand, responses depend on a comparison between two simultaneous auditory signals, increased sound attenuation should lead to smaller responses and/or more opposing responses.

A. Method

1. Participants

49 native speakers of Dutch participated in the experiment in exchange for course credit. All participants were students at Ghent University (41 female and 8 male, mean age = 19.4). None of them had any history of speech, hearing, or language impairments. The study was approved by the ethics committee of the Ghent University faculty of psychology and educational sciences.

2. Procedure

Participants were fitted with a pair of headphones and a head-mounted microphone. On each trial, the appearance of the letters “EE” (pronounced in Dutch as [e]) on a laptop 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 auditory feedback via the headphones. During each vocalization, pitch was shifted for 200ms by either -25 cents, +25 cents, -100 cents, +100 cents, or 0 cents. This happened three times during every vocalization. The addition of 0 cents shifts (null shifts) has two advantages. First, they allowed us to represent responses to pitch shifts not just as deviations from a pre-shift baseline pitch as in previous studies (Bauer and Larson, 2003; Larson et al., 2007; Liu et al., 2012; Liu and Larson, 2007), but also as deviations from “responses” to a null shifts. in this way, a constant pitch drift common to pitch contours in all conditions cannot affect estimations of response direction and magnitude. Second, the presence of null shifts means it was not predictable for participants how many pitch shifts would occur within one vocalization, thus avoiding any anticipation effects. The shifts were separated from each other and from speech onset by a jittered interval of 600 to 800ms. The pitch shifts were randomized within each experimental block in such a way that each set of two consecutive trials contained all four shifts and two null shifts. An experimental block consisted of 80 trials and thus of 240 shifts, including 40 shifts of each perturbation type as well as 80 null shifts. Before each block, participants produced 10 practice vocalizations to get acquainted with the task and the sound of their voice played via the headphones. After each experimental block, participants got a short break during which they changed the headphones. The order of headphones was counterbalanced across all participants.

3. Equipment

Three pairs of headphones were used: the BeyerDynamic DT 770 Pro (hereafter BD), the custom-built headphones (pair nr. 3, hereafter CB), and the Etymotic ER-3C insert earphones (hereafter the ER). Speech was recorded with a head-mounted microphone (DPA 4088-B) 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 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, which connected 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. All signals were stored at a 44.1kHz sampling rate.

In accordance with previous studies, the volume of the auditory feedback was set 10dB above the signal picked up by the microphone (Behroozmand et al., 2014; Hawco et al., 2009; Liu et al., 2011). Any volume differences between headphones were compensated for by adjusting the output gain on the Eclipse Eventide processor. The output gain values used were -16dB, -5dB and -1dB for the CB, the ER and the BD, respectively. These were determined beforehand during a session in which the output volume of each headphone pair was measured with an oscilloscope. The output gain on the Eclipse Eventide processor was adjusted such that all headphones would show a 10dB increase compared to the input volume at the microphone. The delay between microphone input and the auditory feedback output was on average 14.3ms (SD = 5.3ms).

4. Analysis

All data and analysis scripts are publicly available via https://osf.io/vm84u/. The data from three participants was not further analysed, because the ER insert earphones did not fit well and thus could have led to a different feedback volume compared to the other headphones. For one of these three participants, the ER earphones fell out during the experiment. The other two participants reported after the experiment that they felt like the earphones were about to fall out, and had difficulty fitting the earphones before the experiment. For the remaining participants, 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 (200ms before and 700ms after shift onset), the data was included in the analysis.

A pitch estimation algorithm based on autocorrelation in Praat (Boersma and 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 the 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, about 77.3 epochs (i.e., about 10.7 % of the maximum of 720 epochs) were discarded per participant. This includes epochs that displayed pitch tracking errors as well as epochs containing a mistimed pitch shift. The maximal number of epochs discarded for a single participant was 264, with only four participants having more than 200 discarded epochs. Undefined stretches in remaining epochs' pitch contour were linearly interpolated from neighbouring samples.

For each participant, headphones, and perturbation condition, the average pitch response contour was calculated by averaging across epochs, as in previous studies with this paradigm (Bauer and Larson, 2003; Larson et al., 2007; Liu et al., 2012; Liu and Larson, 2007). For each participant, only conditions (i.e., a specific headphones by perturbation combination) in which there were at least 20 epochs were included in further analysis. This resulted in the rejection of the data of two additional participants (they each had no condition with over 20 epochs for two of the three headphones) as well as the rejection of data for three conditions in one participant and one condition in another. The resulting 656 average pitch contours were derived from on average 35.2 epochs (ranging from 20 to 40 out of maximally 40) in the non-null perturbation conditions and from 70.2 epochs (ranging from 33 to 80 out of maximally 80) for the null perturbation conditions. To ensure that response magnitude estimations were not affected by gradual drifts, difference contours were calculated for each participant by subtracting the average for the null perturbation from the average of the corresponding non-null perturbations. The sign of the difference contours for the upward perturbations was flipped such that positive values indicate opposing responses while negative values indicate following responses.

For every pair of headphones and perturbation condition, the compensation response magnitude was estimated as the maximal value after 60ms after the perturbation onset.

In addition, we used a response classification method to classify every single epoch as containing either an opposing or a following response. The epochs were classified based on the slope of the pitch contour over the time window of 60 to 260ms after perturbation onset (Franken et al., 2018a). As 60ms is considered the minimal time that is necessary to respond to a pitch shift (Chen et al., 2007; Larson et al., 2001), this is presumably the window containing possible responses to the pitch shift onset but not (yet) responses to the pitch shift offset. If the slope was positive, the response was labelled as an upward response (i.e., an opposing response for downward perturbations and a following response for upward perturbations). The response classification was run on the difference pitch contours, in order to avoid a bias due to an overall pitch drift that is unrelated to the specific condition’s pitch perturbation.

5. Statistical inference

In order to assess whether pitch shifts led to a general response, each condition's response contour was compared to the response contour for the null shift with the same headphones. This comparison was carried out using a cluster-based permutation test (Maris and Oostenveld, 2007). For every condition, a t-value was calculated at each time sample, and neighbouring 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.

All further statistical tests were carried out in R (R Core Team, 2018). Response magnitudes were entered in linear mixed effects models, with headphone type, perturbation magnitude, and perturbation direction as fixed effects (main effects and all pairwise interactions as well as a three-way interaction) and random intercepts across subjects. The factor headphone type was dummy-coded with the BD as the reference level, while the perturbation direction and the perturbation magnitude were contrast coded. If model convergence allowed it, random slopes across subjects for headphone type, perturbation magnitude, and perturbation direction were added as well (but no random slopes for interaction effects). 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.

The response classification results as either opposing or following responses were entered in a logistic mixed effects model. Reported p values were calculated using Laplace approximation. Omnibus results were derived from type-III Wald χ^2-tests from the 'Anova()'

function in the ‘car’ package (Fox and Weisberg, 2019), while pairwise comparisons were calculated with the ‘emmeans’ package (Lenth, 2019) as before. All mixed effects modelling was performed using the R packages ‘lme4’ (Bates et al., 2015) and ‘lmerTest’ (Kuznetsova et al., 2016).

B. Results

Figure 5 shows the grand average pitch compensation responses, as a function of headphone type, perturbation direction, and perturbation magnitude. These responses show the difference between responses in each condition and the response to a null shift with the same headphones. As expected, in all conditions the grand average pitch contour shows a compensatory response, which starts around 100ms after the perturbation onset and peaks around 250ms after the perturbation onset. At first sight, there seems to be little difference between the responses to the different headphones. Cluster-based permutation tests revealed that for each condition, the response contour differed from the corresponding pitch contour for a null pitch shift (Table 3).

Table 3. Results of the clustered based permutation tests. The reported condition is compared to the corresponding null shift condition in each case. Along with the p-value, the onset time of the largest cluster responsible for the statistical difference is shown.

Headphones	Pitch shift (cents)	Onset Largest Cluster (ms)	*p*
BD	25	101	< .001
BD	-25	206	.018

BD	100	108	< .001
BD	-100	129	< .001
ER	25	162	< .001
ER	-25	152	.022
ER	100	119	< .001
ER	-100	135	.0012
CB	25	130	< .001
CB	-25	159	.0028
CB	100	97	< .001
CB	-100	152	< .001

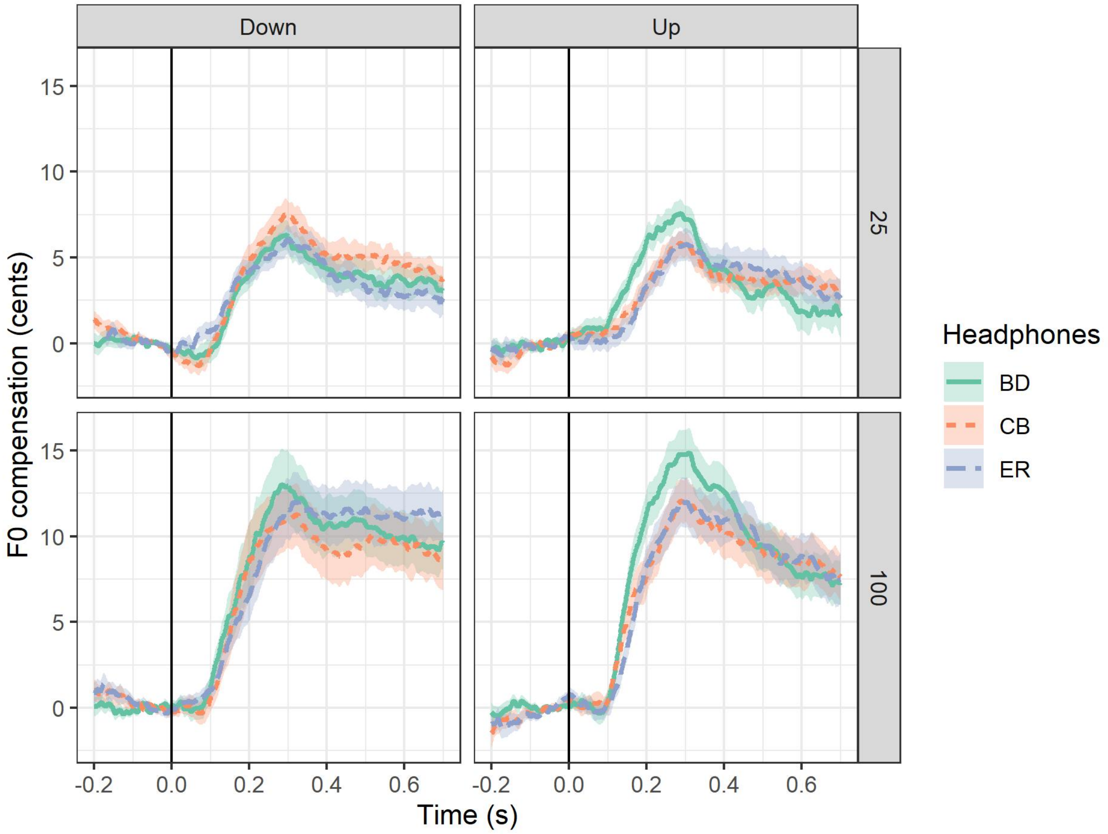

Fig. 5. Grand average pitch compensation contours as a function of Headphones, Perturbation Magnitude, and Perturbation Direction. These contours reflect the difference between the response in each condition and the response to a null shift with the same headphones. The top row displays responses to perturbation with an absolute magnitude of 25 cents, the bottom row displays responses to perturbations with an absolute magnitude of 100 cents. The left column shows responses to downward pitch shifts (i.e., pitch decreases), while the right column shows responses to upward pitch shifts. The sign of the responses to upward pitch shifts was flipped, so positive values indicate an opposing response. Shaded areas around the contours indicate the standard error of the mean. The vertical black lines indicate the perturbation onset.

1. Response Magnitude

The results of a linear mixed effects model of the response magnitude estimates, reported in Table 4, indicate a significant effect of Perturbation Magnitude, showing that responses to 100 cents perturbations were larger than to 25 cents perturbations (contrast = 8.36, SE = 0.67, *t*(451) = 12.46, *p* < .001). Contrary to our expectations, the response magnitude did not vary as a function of headphone type. This suggests that the amount of passive sound attenuation associated with the different headphones did not affect response magnitude. Response magnitude was also not affected by perturbation direction, or any of the 2-way or 3-way interactions between the three factors. The results are visualized in Figure 6.

Table 4. Omnibus fixed effects on the overall response magnitude. The factors pertMag and pertDir refer to perturbation magnitude and perturbation direction respectively. Colons

indicate interaction terms (e.g., Headphones:pertMag refers to the two-way interaction between headphone type and perturbation magnitude).

	SS	df	F	p
Headphones	117.26	2, 451	1.03	.36
pertMag	8797.00	1, 451	155.13	**<.001***
pertDir	2.93	1, 451	0.052	.82
Headphones:pertMag	23.66	2, 451	0.21	.81
Headphones:pertDir	31.04	2, 451	0.27	.76
pertMag:pertDir	4.54	1, 451	0.080	.78
Headphones:pertMag:pertDir	39.75	2, 451	0.35	.70

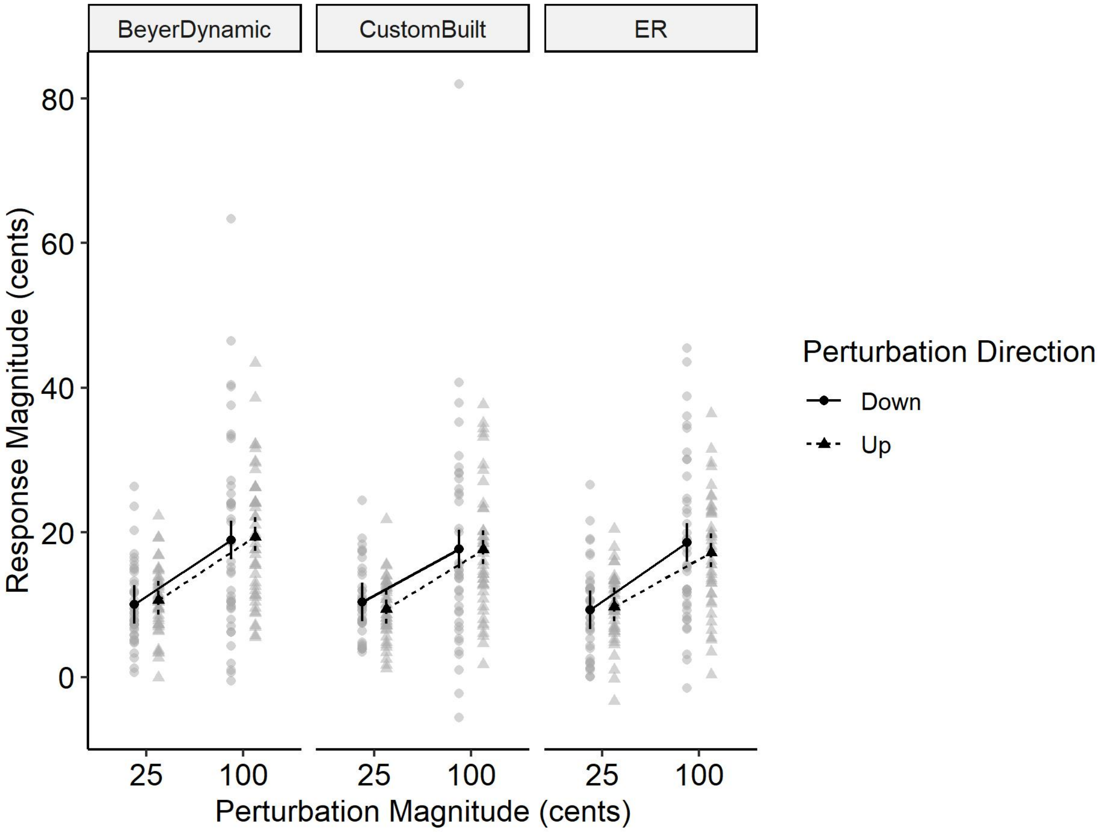

Fig. 6. Response magnitude as a function of Perturbation Magnitude, Perturbation Direction, and Headphones. In grey, the data for individual subjects is plotted. In black, the fitted values from the mixed effects model are plotted. The error bars indicate 95% confidence intervals.

In a second analysis, the response magnitude was also quantified using only the epochs classified as having opposing responses. Again, only the perturbation magnitude affected response magnitude (contrast = 7.45, $t(412) = 11.31$, $p < .001$). None of the other main effects or interactions yielded significant results.

2. Proportion of opposing responses

Next, epochs were classified as containing either an opposing or a following response. Out of a total of 19857 analysed epochs across all participants and conditions (the null shift excluded), 13377 (about 67%) were classified as opposing and 6480 as following. The probability of an opposing response (‘opposing probability’) was modelled as a function of the perturbation magnitude, perturbation direction, and type of headphones in a logistic mixed effects model. The results are visualized in Figure 7 and the omnibus effects are shown in Table 5. The results suggest a main effect of headphones ($\chi^2(2) = 7.39$, $p = .025$), a main effect of perturbation magnitude ($\chi^2(1) = 35.45$, $p < .001$), a marginally significant main effect of perturbation direction ($\chi^2(1) = 3.12$, $p = .077$), as well as significant two-way interactions between these factors. Closer examination of the interaction between headphone type and perturbation magnitude suggested that for 25 cents perturbations, the ER led to a lower opposing probability compared to the BD (est. = 0.18, $z = 2.49$, $p = .034$) as well as compared to the CB (est. = 0.23, $z = 2.84$, $p = .013$). For the 100 cents perturbations, there were no significant pairwise contrasts. A similar pattern is visible for the interaction between headphone type and perturbation direction. For downward perturbations, the ER show a lower opposing probability compared to CB (est. = 0.24, $z = 3.01$, $p = .0074$) and a trend towards a lower opposing probability when compared to BD (est. = 0.17, $z = 2.21$, $p = .070$). For upward perturbations, there are no significant contrasts, although there is a trend of BD showing higher opposing probability compared to either CB (est. = 0.18, $z = 2.23$, $p = .066$) or ER (est. = 0.18, $z = 2.32$, $p = .053$).

Table 5. Omnibus (Type-III Wald χ^2 tests) results for opposing probability.

	χ^2	df	p

(Intercept)	124.50	1	**< .001***
Headphones	7.39	2	**.025***
pertMag	35.45	1	**< .001***
pertDir	3.12	1	.077
Headphones:pertMag	7.92	2	**.019***
Headphones:pertDir	12.82	2	**.0016***
pertMag:pertDir	20.94	1	**< .001***
Headphones:pertMag:pertDir	1.08	2	.58

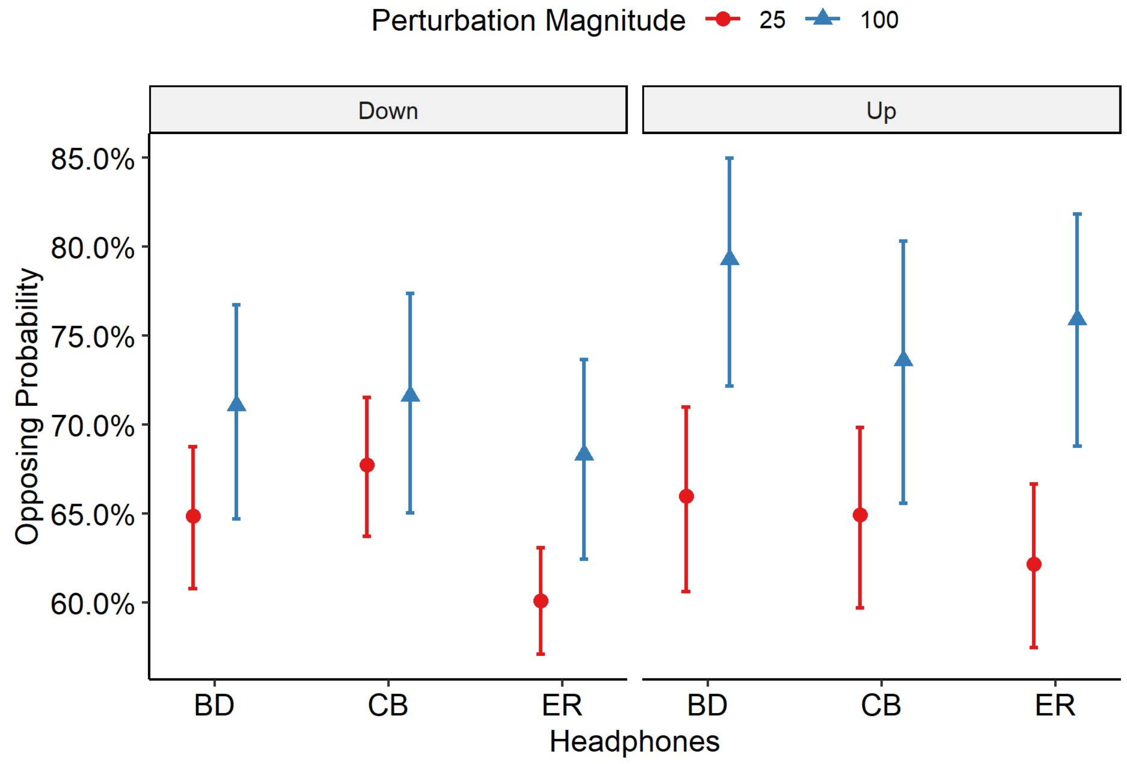

Fig. 7. The probability of opposing responses as a function of Perturbation Magnitude, Perturbation Direction, and Headphones. The error bars reflect the 95% confidence intervals of the model's fixed effect estimates.

In addition, the effect of perturbation magnitude interacted with perturbation direction, suggesting that the difference in probability of opposing responses for 25 cents and 100 cents perturbation was larger for upward (est. = 0.58, z = 7.29, p < .001) than for downward perturbations (est. = 0.28, z = 3.50, p < .001). The three-way interaction between headphone type, perturbation magnitude, and perturbation direction was not significant.

IV. GENERAL DISCUSSION

The current study investigated speakers' responses to pitch perturbations with three different headphones varying in amount of passive sound attenuation. Our main research question was whether there are sound attenuation-related differences in the responses to pitch shifts. If responses to pitch shifts are driven by a comparison between an internal pitch target and perceived auditory feedback, increased passive sound attenuation would make the discrepancy between target and feedback more salient, leading to larger responses for more attenuating headphones (like ER). On the other hand, if responses are driven by a comparison between the manipulated signal and the original feedback leaking through the headphones, increased attenuation would make the discrepancy less salient and thus more attenuating headphones should lead to smaller responses. Similarly, if increased attenuation makes it more likely that the manipulated feedback is considered by the speaker as self-generated, we expect more sound attenuation to lead to more opposing responses. In terms of response magnitudes, there were no differences between headphones, in contrast with our hypotheses. This null result suggests that in terms of response magnitude, it does not matter which of the three headphones was used.

In terms of the response type (i.e., proportion of following vs. opposing responses), the current analysis revealed no clear overall association between sound attenuation and response type, although there was an interaction of headphone type with both perturbation direction and perturbation magnitude. This pattern of results, although not consistent, suggests that the type of headphones does play a role in this paradigm, although it is hard to pinpoint what role precisely. Specific contrasts showed that response types were only affected by headphone type in some conditions. Although these contrasts may tentatively suggest that higher attenuation (as in ER) is associated with fewer opposing responses, this should be interpreted with caution since a number of the examined contrasts are not strictly significant, and it is unclear how the interactions with perturbation direction and magnitude should be interpreted. In addition, the results of Experiment 1 suggested that the ER show a somewhat different frequency response compared to the circumaural headphones, suggesting that differences that only affect the ER without a difference between CB and BD could be driven by either sound attenuation or the different frequency responses. If, however, future work would corroborate a link between more sound attenuation and less opposing responses, this suggests that passive sound attenuation affects response type but not response magnitude. This is in contrast with our hypotheses. Previous studies showed that response magnitude varies with perturbation magnitude (Chen et al., 2007; Hawco et al., 2009; Liu and Larson, 2007), suggesting that response magnitude can be treated as an index of the conflict introduced by the feedback perturbation. While the causes of following responses are unclear, several authors have proposed that one contributing factor may be that participants treat the feedback signal itself as a referent (Hain et al., 2000; Patel et al., 2014), rather than as self-generated auditory feedback. If so, the current (tentative) results suggest that the sound attenuation of the headphones affects the participants' source monitoring of the auditory signal, but not the magnitude of the feedback mismatch itself.

Why would sound attenuation affect response type but not response magnitude? The auditory feedback in the current study was set at 10dBA louder than the signal picked by the microphone, which leads to a quite loud feedback signal. In fact, some of the participants in the current study spontaneously noted that the feedback was very loud. Increasing the loudness of the feedback signal, like many studies do to try and drown out bone-conducted auditory feedback (Behroozmand et al., 2014; Chen et al., 2007; Liu et al., 2011), may create an atypical situation as speakers do not usually hear themselves so loud. If the volume-induced unnaturalness of this feedback signal is exacerbated by the high passive sound attenuation in the ER, it could lead participants to treat the signal as an external referent and to follow the feedback. Just as very large perturbation magnitudes are considered to lead to following because they are unlikely to be self-generated, very high sound attenuation combined with louder than usual auditory feedback could lead to an intrusive auditory signal, which is unlikely to be self-generated. However, the weak statistical evidence in the current study as well as the absence of an effect on the magnitude of the (opposing) responses shows that additional work is necessary to disentangle these possibilities.

If the differences between headphones in the current study turn out to be false positives, in line with the absence of a clear overall association between attenuation and response magnitude and types, this would suggest that the differences in passive sound attenuation in the current study did not play a significant role in responses to pitch-shifted feedback, as suggested also by the null effect for the response magnitude. This suggests that compensatory responses across different pitch perturbation studies should be comparable regardless of the headphones used. Although it is possible that varying sound attenuation could have different results if other speech features (e.g., formant values) were manipulated, the current result seems to be in line with a previous study perturbing the first formant (Mitsuya and Purcell, 2016). However, we should be cautious with this conclusion given that there are possible alternative explanations for the absence of an effect of sound attenuation. This result may indicate that the passive sound attenuation of all three headphones in the current study is either good enough to make the manipulated feedback dominant over any leaking original feedback or that participants simply treat only the loudest auditory input as their feedback signal, but it is similarly possible that the differences in attenuation are not large enough, and that therefore non-manipulated auditory feedback plays a similar role in all three headphones. We speculate that this may be caused, in part, by bone-conducted auditory feedback which is potentially not affected by the attenuation properties of the different headphones[3]. Thus, strictly speaking, the current results are not able to distinguish between the dominant hypothesis suggesting that responses are dependent on a comparison between an internal pitch representation and the manipulated feedback, and the alternative hypothesis where responses are dependent on the contrast between the manipulated and non-manipulated auditory feedback.

For future studies, an interesting way to address this issue is to mask bone-conducted auditory feedback in order to limit its role and isolate air-conducted feedback that would be affected by headphones’ sound attenuation. One could attempt this by playing speech-shaped noise through bone conduction headphones while manipulated feedback is played through normal headphones as in the current experiment. Another way forward may be to take more control over the relative level of the normal and manipulated feedback signals by playing both normal and manipulated feedback through a single pair of headphones while varying the relative levels of both signals. In most altered auditory feedback experiments, it is common to amplify the auditory feedback (as in the current study), in an attempt to make it more salient than potentially conflicting feedback signals. An experiment that compares different relative loudness levels of non-manipulated and manipulated feedback would yield more insight into the role of the relative weighting of conflicting feedback signals.

As expected, speakers in the current study show stronger compensation responses to larger perturbations, in line with previous studies (Chen et al., 2007; Hafke, 2008; Hawco et al., 2009; Liu and Larson, 2007), although others have failed to find such an effect (Burnett et al., 1998; Liu et al., 2010b). Interestingly, some studies have shown that this relationship between perturbation magnitude and response magnitude holds only for relatively small perturbations (i.e., up to 200/250 cents), with the compensation response decreasing again for larger responses (Behroozmand et al., 2012; Scheerer et al., 2013), because very large perturbations are unlikely to be considered to be self-generated by the speaker. In addition, we find that 100 cents perturbations in the current study led to a higher probability of opposing responses compared to 25 cents perturbations. Although this does not speak to the influence of leaking non-manipulated auditory feedback, this result is in contrast with some previous studies finding more following responses with larger magnitudes (Burnett et al., 1998; Liu et al., 2010a, 2011, 2010b). It is important to note here that the perturbation magnitudes used in these previous studies were larger than in the present study: for example, most of these studies (Liu et al., 2010a, 2010b) found more following responses to 200 or 500 cents perturbations compared to smaller (50 cents and 100 cents) perturbations. As with the findings of differences in response magnitude, based on the findings in the current and in previous studies, we propose that the response type (following vs. opposing) may also show an (inverted) U-shaped relationship with the perturbation magnitude. On the one hand, literature suggests that following responses occur more frequently with very large pitch shifts (200, 500 cents), which may be due to large perturbations being less likely to be recognized as self-produced by the speaker and therefore participants follow it as an external pitch referent. This is in line with suggestions that following responses are observed when the speaker does not consider the presented auditory feedback signal as self-produced speech (Hain et al., 2000; Patel et al., 2014). On the other hand, while very small perturbations (e.g., 25 cents in the current study) are highly likely to be self-generated, they lead to fewer opposing responses as the shifts are less salient compared to slightly larger shifts that are still considered to be self-generated (e.g., 100 cents). In the same vein, a 25 cents shift leads to a smaller compensation response than a 100 cents shift because it is less salient, while a 500 cents shift leads to a smaller response compared to 100 cents shifts because it is no longer considered to be self-generated.

In addition, it is important to note that most of the previous studies identified the response type (following or opposing) at the average level: epochs were averaged for every condition and participant, and it was identified whether this average response was either following or opposing. Given that recent studies suggest that speakers generally both oppose and follow the pitch shift even within the same condition (Behroozmand et al., 2012; Franken et al., 2018a), the current study classified responses at the single epoch level. Behroozmand et al. (2012) did the same but found no effect of perturbation magnitude on the amount of following/opposing responses (they used perturbations of 50, 100, and 200 cents). Given the variability at the single epoch level, it may be the case that correct response classification is harder for 25 cents shifts, as the response magnitudes are smaller and therefore have a lower signal-to-noise ratio. A more precise characterization of the effect of perturbation magnitude on the frequency of opposing and following responses deserves further investigation.

With respect to the probability of opposing responses, the current results showed an interaction between perturbation magnitude and perturbation direction, with a stronger effect of perturbation magnitude on opposing probability in the upward shifts compared to the downward shifts. To our knowledge, this is the first paper reporting that the response type may vary as a function of direction, but previous studies have shown directionality effects on response magnitude. The current results seem in contrast with some studies showing larger responses to downward shifts compared to upward shifts (Liu et al., 2011; Liu and Larson, 2007; Sturgeon et al., 2015), while others have found no effect of perturbation direction (Larson et al., 2001, 2008). In the current study we find no effect of perturbation direction on response magnitude, but only on opposing probability. Instead of a directionality effect, this could also suggest an overall bias, in our sample, for downward responses, which would be opposing in response to upward shifts and following in response to downward shifts. We suggest further investigation is needed to investigate the effect of the direction on pitch response types.

Overall, the current results suggest that passive sound attenuation has no effect on response magnitudes to unexpected pitch shifts in online auditory feedback. In addition, sound attenuation did also not have a clear effect on the response type. While response type may be affected by multiple factors, including pitch fluctuations before the perturbation onset (Franken et al., 2018a) and properties of the pitch manipulation (Burnett et al., 1998; Liu et al., 2010b), we suggest that in the current study it may be treated as an index of source monitoring. In other words, response type could reflect whether participants attribute the pitch shift to their own production or to an external source. Although we should be cautious to interpret the weak evidence in the current study, we propose that it is important to take into account that non-manipulated auditory feedback may not be completely masked in pitch-shift studies. Responses to pitch-shifted feedback are not only driven by the mismatch between an internal speech target and the manipulated auditory signal, but potentially also by the source attributed to the auditory signal by the speaker. We suggest that it would be interesting for future studies to measure both the response magnitude as well as the response types at an epoch by epoch level. In addition, we have suggested that both response magnitude and response type show an inverted U-shaped relationship with pitch-shift magnitude: for small perturbations, which are likely to be treated as self-generated, larger perturbations lead to larger responses and more opposing responses. Other studies have suggested, in addition, that very large perturbations lead to a decrease in response magnitude and an increase in following responses. Both error-monitoring in speech production as well as source monitoring are functions that have been associated with auditory feedback processing previously (Hain et al., 2000; Korzyukov et al., 2017; Subramaniam et al., 2018). In future studies, it will be important to further investigate the interplay between these two processes.

ACKNOWLEDGEMENTS

We are grateful to Pascal Mestdagh for his help with the acoustic measurements and to Nicolas Bourguignon for fruitful discussions. MKF was supported by the Special Research Fund (BOF) at Ghent University. AL was supported by the Swedish Research Council (Grant nr. 2015-00446) and the Crafoord Foundation (Grant nr. 20170734). LH and PH were supported by the Swedish Research Council (Grant nr. 2014-1371) and the Crafoord

Foundation. TW was supported by a European Research Council starting grant (ERC-StG-678120 ('RobSpear').

FOOTNOTES

[1] The headphones were conceived and constructed by the last author.

[2] See supplementary material at [URL will be inserted by AIP] for details on the construction of the custom-built pair of headphones.

[3] With current technology, it is difficult to know for sure whether bone-conducted feedback is masked by manipulated auditory feedback or not.

REFERENCES

Alain, C. (**2007**). "Breaking the wave: Effects of attention and learning on concurrent sound perception," Hear. Res., **229**, 225–236. doi:10.1016/j.heares.2007.01.011

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting Linear Mixed-Effects Models Using lme4," J. Stat. Softw., , doi: 10.18637/jss.v067.i01. doi:10.18637/jss.v067.i01

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am., **114**, 1048. doi:10.1121/1.1592161

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

Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C. R. (**2012**). "Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control," J. Acoust. Soc. Am., **132**, 2468–77. doi:10.1121/1.4746984

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., and Greenlee, J. D. W. (**2015**). "Sensory-motor networks involved in speech production and motor control: an fMRI study," Neuroimage, **109**, 418–28. doi:10.1016/j.neuroimage.2015.01.040

Boersma, P., and Weenink, D. (**2017**). "Praat: doing phonetics by computer [Computer Program].," Retrieved from http://www.praat.org

Breebaart, J. (**2017**). "No correlation between headphone frequency response and retail price," J. Acoust. Soc. Am., , doi: 10.1121/1.4984044. doi:10.1121/1.4984044

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am., **103**, 3153–3161. doi:10.1121/1.423073

Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (**2010**). "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization," J. Acoust. Soc. Am., **128**, 2033–48. doi:10.1121/1.3479539

Casserly, E. D. (**2011**). "Speaker compensation for local perturbation of fricative acoustic feedback," J. Acoust. Soc. Am., **129**, 2181–2190. doi:10.1121/1.3552883

Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F[sub 0] responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am., **121**, 1157. doi:10.1121/1.2404624

Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci., **1**, 327. Retrieved from http://web.mit.edu/hst.723/www/ThemePapers/ASA/Darwin97.pdf

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). “Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers,” J. Acoust. Soc. Am., **114**, 2913. doi:10.1121/1.1616924

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). “ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment,” Int. J. Audiol., **40**, 148–157. doi:10.3109/00206090109073110

Elman, J. L. (**1981**). “Effects of frequency-shifted feedback on the pitch of vocal productions,” J. Acoust. Soc. Am., **70**, 45. doi:10.1121/1.386580

Epstein, A., Giolas, T. G., and Owens, E. (**1968**). “Familiarity and Intelligibility of Monosyllabic Word Lists,” J. Speech Hear. Res., **11**, 435–438. doi:10.1044/jshr.1102.435

Flagmeier, S. G., Ray, K. L., Parkinson, A. L., Li, K., Vargas, R., Price, L. R., Laird, A. R., et al. (**2014**). “The neural changes in connectivity of the voice network during voice pitch perturbation,” Brain Lang., **132**, 7–13. doi:10.1016/j.bandl.2014.02.001

Fox, J., and Weisberg, S. (**2019**). *An {R} Companion to Applied Regression*, Sage, Thousand Oaks, CA, Third edit. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., and Eisner, F. (**2018**). “Opposing and following responses in sensorimotor speech control: Why responses go both ways,” Psychon. Bull. Rev., **25**, 1458–1467. doi:10.3758/s13423-018-1494-x

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., and Eisner, F. (**2019**). “Consistency influences altered auditory feedback processing,” Q. J. Exp. Psychol., **72**, 2371–2379. doi:10.1177/1747021819838939

Franken, M. K., Eisner, F., Acheson, D. J., McQueen, J. M., Hagoort, P., and Schoffelen, J. (**2018**). “Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production," Neuroimage, **179**, 326–336. doi:10.1016/j.neuroimage.2018.06.061

Guenther, F. H. (**2016**). *Neural Control of Speech*, The MIT Press, Cambridge, MA, 424 pages.

Hafke, H. Z. (**2008**). "Nonconscious control of fundamental voice frequency," J. Acoust. Soc. Am., **123**, 273–8. doi:10.1121/1.2817357

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res., **130**, 133–141. doi:10.1007/s002219900237

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

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science (80-.)., **279**, 1213–1216. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9469813

Houde, J. F., and Nagarajan, S. S. (**2011**). "Speech production as state feedback control," Front. Hum. Neurosci., , doi: 10.3389/fnhum.2011.00082. doi:10.3389/fnhum.2011.00082

Johnston, J. D. (**1988**). "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE J. Sel. areas Commun., Retrieved from https://pdfs.semanticscholar.org/c147/3395b8cc9714930507bde86c89d9d931f6ea.pdf. Retrieved from https://pdfs.semanticscholar.org/c147/3395b8cc9714930507bde86c89d9d931f6ea.pdf

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am., **108**, 1246. doi:10.1121/1.1288414

Keough, D., and Jones, J. A. (**2009**). "The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing," J. Acoust. Soc. Am., **126**, 837–46. doi:10.1121/1.3158600

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

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2016**). "lmerTest: Tests in Linear Mixed Effects Models R Package version 20-33.," Retrieved from https://cran.r-project.org/package=lmerTest

Lametti, D. R., Nasir, S. M., and Ostry, D. J. (**2012**). "Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback," J. Neurosci., **32**, 9351–9358. doi:10.1523/JNEUROSCI.0404-12.2012

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., and Ostry, D. J. (**2014**). "Plasticity in the Human Speech Motor System Drives Changes in Speech Perception," J. Neurosci., **34**, 10339–10346. doi:10.1523/JNEUROSCI.0108-14.2014

Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (**2008**). "Interactions between auditory and somatosensory feedback for voice F0 control," Exp. brain Res., **187**, 613–21. doi:10.1007/s00221-008-1330-z

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparison of voice F[sub 0] responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am., **110**, 2845. doi:10.1121/1.1417527

Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F[sub 0] and amplitude control," J. Acoust. Soc.

Am., **121**, 2862. doi:10.1121/1.2715657

Lenth, R. (**2019**). “emmeans: Estimated Marginal Means, aka Least-Squares Means.,” Retrieved from https://cran.r-project.org/package=emmeans

Li, W., Chen, Z., Liu, P., Zhang, B., Huang, D., and Liu, H. (**2013**). “Neurophysiological evidence of differential mechanisms involved in producing opposing and following responses to altered auditory feedback,” Clin. Neurophysiol., **124**, 2161–71. doi:10.1016/j.clinph.2013.04.340

Lind, A., Hall, L., Breidegard, B., Balkenius, C., and Johansson, P. (**2014**). “Auditory feedback of one’s own voice is used for high-level semantic monitoring: the ‘self-comprehension’ hypothesis,” Front. Hum. Neurosci., , doi: 10.3389/fnhum.2014.00166. doi:10.3389/fnhum.2014.00166

Liu, H., and Larson, C. R. (**2007**). “Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex,” J. Acoust. Soc. Am., **122**, 3671–7. doi:10.1121/1.2800254

Liu, H., Meshman, M., Behroozmand, R., and Larson, C. R. (**2011**). “Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback,” Clin. Neurophysiol., **122**, 951–7. doi:10.1016/j.clinph.2010.08.010

Liu, H., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., and Huang, D. (**2010**). “Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations,” J. Acoust. Soc. Am., **128**, 3739–3746. doi:10.1121/1.3500675

Liu, H., Wang, E. Q., Metman, L. V., and Larson, C. R. (**2012**). “Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson’s disease,” PLoS One, **7**, e33629. doi:10.1371/journal.pone.0033629

Liu, P., Chen, Z., Larson, C. R., Huang, D., and Liu, H. (**2010**). “Auditory feedback control of voice fundamental frequency in school children," J. Acoust. Soc. Am., **128**, 1306. doi:10.1121/1.3467773

Maris, E., and Oostenveld, R. (**2007**). "Nonparametric statistical testing of EEG- and MEG-data," J. Neurosci. Methods, **164**, 177–190. doi:DOI 10.1016/j.jneumeth.2007.03.024

Mitsuya, T., and Purcell, D. W. (**2016**). "Occlusion effect on compensatory formant production and voice amplitude in response to real-time perturbation," Artic. J. Acoust. Soc. Am., , doi: 10.1121/1.4968539. doi:10.1121/1.4968539

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., and Ivry, R. B. (**2017**). "Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration," J. Neurosci., **37**, 9249–9258. doi:10.1523/JNEUROSCI.3363-16.2017

Patel, S., Nishimura, C., Lodhavia, A., Korzyukov, O., Parkinson, A., Robin, D. a, and Larson, C. R. (**2014**). "Understanding the mechanisms underlying voluntary responses to pitch-shifted auditory feedback," J. Acoust. Soc. Am., **135**, 3036. doi:10.1121/1.4870490

Pörschmann, C. (**2000**). "Influences of Bone Conduction and Air Conduction on the Sound of One's Own Voice," Acust. Acta Acust., **86**, 1038–1045. Retrieved from www.ingentaconnect.com/content/dav/aaua/2000/00000086/00000006/art00018

Puckette, M. (**1996**). "Pure data," Proc. Int. Comput. Music Conf., International Computer Music Association, San Francisco, CA, 269–272.

Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am., **120**, 966–977. doi:10.1121/1.2217714

R Core Team (**2018**). "R: a language and environment for statistical computing.," Retrieved from http://www.r-project.org

Scheerer, N. E., Behich, J., Liu, H., and Jones, J. A. (**2013**). "ERP correlates of the magnitude of pitch errors detected in the human voice," Neuroscience, **240**, 176–85. doi:10.1016/j.neuroscience.2013.02.054

Schuerman, W. L., Nagarajan, S., McQueen, J. M., and Houde, J. (**2017**). "Sensorimotor adaptation affects perceptual compensation for coarticulation," J. Acoust. Soc. Am., **141**, 2693–2704. doi:10.1121/1.4979791

Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (**2009**). "Perceptual recalibration of speech sounds following speech motor learning," J Acoust Soc Am, **125**, 1103–1113. doi:10.1121/1.3058638

Sturgeon, B. A., Hubbard, R. J., Schmidt, S. A., and Loucks, T. M. (**2015**). "High F0 and musicianship make a difference: Pitch-shift responses across the vocal range," J. Phon., **51**, 70–81. doi:10.1016/j.wocn.2014.12.001

Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S. S., and Houde, J. F. (**2018**). "Reality Monitoring and Feedback Control of Speech Production Are Related Through Self-Agency," Front. Hum. Neurosci., **12**, 82. doi:10.3389/fnhum.2018.00082

Warren, R. M., Riener, K. R., Bashford, J. A., and Brubaker, B. S. (**1995**). *Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits* Percept. Psychophys.,Vol. 57, 175–182 pages. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03206503.pdf