The influence of coarticulatory and phonemic relations on individual compensatory formant production

Eugen Klein[a] and Jana Brunner
Institut für Deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin, Berlin, Germany

Phil Hoole
Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München, München, Germany

(Received 21 November 2018; revised 29 July 2019; accepted 31 July 2019; published online 21 August 2019)

Previous auditory perturbation studies have shown that speakers are able to simultaneously use multiple compensatory strategies to produce a certain acoustic target. In the case of formant perturbation, these findings were obtained examining the compensatory production for low vowels /ɛ/ and /æ/. This raises some controversy as more recent research suggests that the contribution of the somatosensory feedback to the production of vowels might differ across phonemes. In particular, the compensatory magnitude to auditory perturbations is expected to be weaker for high vowels compared to low vowels since the former are characterized by larger linguopalatal contact. To investigate this hypothesis, this paper conducted a bidirectional auditory perturbation study in which *F*2 of the high central vowel /ɨ/ was perturbed in opposing directions depending on the preceding consonant (alveolar vs velar). The consonants were chosen such that speakers' usual coarticulatory patterns were either compatible or incompatible with the required compensatory strategy. The results demonstrate that speakers were able to compensate for applied perturbations even if speakers' compensatory movements resulted in unusual coarticulatory configurations. However, the results also suggest that individual compensatory patterns were influenced by additional perceptual factors attributable to the phonemic space surrounding the target vowel /ɨ/.
 https://doi.org/10.1121/1.5122788

[ZZ] Pages: 1265–1278

I. INTRODUCTION

During speech production, speakers push a controlled air stream past their vocal folds and through a vocal tract configuration formed by a set of articulators (tongue, lips, jaw, upper incisors, etc.), which ultimately results in a certain acoustic output. In order to understand what articulatory and/or acoustic parameters speakers try to control for during this process, it proved empirically successful to perturb speakers' sensory feedback and to observe whether and how they correct for the induced errors.

The auditory feedback, i.e., the perception of one's own voice, is the most obvious type of sensory feedback present during speech production. Due to technical innovation, in more recent years it has become possible to alter, in near real-time, specific acoustic parameters such as fundamental frequency (*f*0; Jones and Munhall, 2000), vowel formants (*F*1 and/or *F*2; Houde and Jordan, 1998; Purcell and Munhall, 2006; Villacorta *et al.*, 2007), and frication noise (Shiller *et al.*, 2009) in speakers' auditory feedback.

During an auditory perturbation study, speakers are usually asked to repeatedly produce a short word or syllable while they are hearing themselves over headphones. When the perturbation is applied, such that, for instance, *F*1 is increased in the auditory feedback, speakers begin to perceive the experimental stimulus differently from how they actually produce it. For instance, if the vowel /ɛ/ is perturbed in this way, it starts to resemble a /æ/. In response, speakers typically decrease *F*1 in their produced speech in order to restore their percept of the intended word. That means that the produced *F*1 may become comparable to that of a /ɪ/. This compensatory response is generalizable across most investigated acoustic parameters and demonstrates that under auditory perturbation speakers try to maintain their auditory target by adjusting their articulatory movements.

The magnitude of the compensatory response is known to be influenced by some perceptual processes. For instance, Villacorta *et al.* (2007) demonstrated that speakers' individual auditory acuity scores with regard to *F*1 frequency significantly correlated with the magnitude of their compensatory response during *F*1 perturbation. Furthermore, the findings of Niziolek and Guenther (2013) suggest that the magnitude of the compensatory response does not depend purely on acoustic distance between produced and perturbed targets, but can become much larger when the perturbation results in a phonemic category change of the perturbed vowel compared to only sub-phonemic changes. This is consistent with findings by Reilly and Dougherty (2013), who showed that speakers react less strongly to *F*1 perturbations if *F*1 constitutes a less important perceptual cue for the identification and discrimination of the perturbed vowel.

Beside the auditory feedback, which is essential for controlling the acoustic target of the speech signal, the somatosensory feedback was also shown to play an important role during speech production. Particularly, the experiments by Tremblay *et al.* (2003) and Nasir and Ostry (2006) demonstrated that

[a]Electronic mail: eugen.klein@hu-berlin.de

speakers compensate for jaw movement perturbations delivered by a robotic arm. These authors found that, although the jaw perturbation did not have any measurable effect on the acoustic outcome of speakers' articulation, approximately 50% of them compensated for it, which suggests that speakers were aware of somatosensory errors and actively tried to correct these. Furthermore, the authors did not find the same compensatory effects on trials where speakers produced opening non-speech jaw movements.

Since the role of auditory and somatosensory feedback has been investigated in most studies separately, it is not completely clear how both feedback signals are incorporated during speech production. More recently, Lametti *et al.* (2012) hypothesized that speakers might exhibit individual preferences regarding the sensory feedback channel they predominantly employ to monitor their own speech production. In their study, the authors investigated participants' responses to a simultaneous somatosensory jaw and auditory $F1$ perturbation. The results from Lametti *et al.* revealed a minor negative correlation across participants between the amount of observed somatosensory and auditory compensation. This means that speakers who changed their jaw position, compensating for the somatosensory perturbation, did not significantly change their $F1$ during auditory perturbation and vice versa.

Unlike the Lametti *et al.* (2012) study, in which $F1$ was perturbed upwards causing an auditory error compatible with the simultaneously applied jaw opening perturbation, perturbations applied to either auditory or somatosensory feedback signal may induce incompatible information in the speech production system. Particularly, while auditory feedback might signal an error, somatosensory feedback might indicate that the appropriate target was achieved. Some authors suggested that such incongruence between feedback signals could be a potential reason for partial compensations observed during formant perturbation (cf. MacDonald *et al.*, 2010; Katseff *et al.*, 2012).

The importance of congruency between specific auditory and somatosensory targets was, however, questioned by some empirical findings. For instance, Rochet-Capellan and Ostry (2011) demonstrated that speakers can simultaneously use multiple articulatory configurations to produce one vowel. To show this, the authors let their speakers repeatedly produce the words "head," "bed," and "ted," while $F1$ in the vowel /ɛ/ was perturbed in opposing directions in two stimuli and remained unchanged in a control stimulus. On average, speakers were able to consistently compensate for the opposing $F1$ perturbations as well as to keep their $F1$ unchanged in the control stimulus. In other words, in that particular case, speakers employed three different articulatory configurations to produce the vowel /ɛ/ as long as their auditory feedback suggested that they achieved the $F1$ value corresponding to their usual acoustic target of this vowel. Consistent with these findings, Feng *et al.* (2011) demonstrated that speakers compensated for the closing perturbation of the jaw during the production of vowels /ɛ/ and /æ/ only when it resulted in a measurable $F1$ decrease. When, at the same time, $F1$ was increased in participants' auditory feedback to match their intended acoustic output, they no longer compensated for the jaw perturbation.

The findings by Rochet-Capellan and Ostry (2011) and Feng *et al.* (2011) are, however, limited to $F1$ perturbation in low vowels /ɛ/ and /æ/. Any generalizations based on these results may be difficult as more recent auditory perturbation research suggests that the contribution of the somatosensory feedback to the production of vowels might differ across different phonemes. In particular, the compensatory magnitude to auditory perturbations is expected to be weaker for high vowels such as /i/ compared to low vowels such as /ɛ/ and /æ/ since the former are characterized by a larger physical contact between active (tongue) and passive (hard palate) articulators (cf. Mitsuya *et al.*, 2015). In other words, the incongruence between the auditory and somatosensory feedback signals might play a more important role for high vowels compared to low vowels. Therefore, to evaluate this hypothesis, in our current study we investigated compensatory effects associated with opposing auditory perturbation applied to a high vowel.

In contrast to Rochet-Capellan and Ostry (2011) and Mitsuya *et al.* (2015), we decided to perturb $F2$ frequency which, roughly speaking, is an indicator of the horizontal tongue displacement. In combination with the perturbation of a high vowel, this change allowed us to evaluate our hypothesis under even more prominent somatosensory feedback conditions since, in order to compensate for the applied perturbation, our speakers were required to always retain the linguopalatal contact in the target vowel while moving its constriction location along the anterior-posterior axis.

As basis for our experimental design, we chose the Russian phoneme space since it includes the high central unrounded vowel /ɨ/, which is enclosed by the two high vowels /i/ and /u/. During the study, while participants repeatedly produced the target vowel /ɨ/ as part of CV syllables, its $F2$ was perturbed in opposing directions depending on the preceding consonant (/d/ or /g/). The bidirectional perturbation of $F2$ was intended to encourage participants to use two different compensatory strategies to produce the perturbed vowel, and the two different consonantal contexts (alveolar vs velar) were chosen in such a way that the required compensatory directions were either compatible or incompatible with the usual coarticulatory relation between /dɨ/ and /gɨ/. The interaction between the place of articulation (alveolar vs velar) and the $F2$ perturbation direction (upward vs downward) was evenly counterbalanced between all participants. This resulted in two different coarticulatory configurations that were tested across two experimental groups.

Due to coarticulatory effects, speakers' $F2$ was expected to be higher in /dɨ/ compared to /gɨ/ before any compensatory response occurred. In one group, $F2$ was decreased in /dɨ/ and increased in /gɨ/ such that the initial $F2$ values of the two syllables were expected to drift apart over the duration of the experiment but to remain in their initial coarticulatory relation (/dɨ/ > /gɨ/). In the other group, the perturbation direction was swapped for both syllables, putatively preventing effective compensation as it was counteracted by coarticulatory effects, and the initial $F2$ values had to intersect for the two syllables over the duration of the experiment. This means that participants in the second group had to produce the two experimental

syllables with an unusual coarticulatory pattern where *F*2 for /gɨ/ would be higher compared to /dɨ/.

One phonemic idiosyncrasy of Russian high vowels, which was expected to have a further influence on the magnitude of speakers' compensatory responses, is the fact that while /i/ appears only after palatalized consonants, both /ɨ/ and /u/ follow only non-palatalized ones (cf. Bolla, 1981). The acoustic feature which is most strongly associated with palatalized consonants is the height of the *F*2 frequency at the beginning of the following vowel, which has highest values for /i/ compared to /ɨ/ and /u/. The height of *F*2 is so dominant for Russian speakers as perceptual cue to palatalization that even cross-spliced syllables containing non-palatalized consonants and vowels with high initial *F*2 frequency are mostly perceived as palatalized (Bondarko, 2005). Since the difference in *F*2 values between /i/ and /ɨ/ is on average substantially smaller (about 300 Hz) compared to /ɨ/ and /u/ (about 1000 Hz), it seems reasonable that, under equivalent amount of upward and downward *F*2 perturbation, participants should more readily classify instances of /ɨ/ perturbed towards /i/ as phonemic errors of palatalization compared to the perturbation of /ɨ/ towards /u/, which does not induce a change of the phonemic status of the perceived vowel. Consequently, considering the findings by Niziolek and Guenther (2013), speakers' compensatory responses should be on average stronger for the upward perturbation compared to the downward perturbation. However, since statistical analyses of the compensatory behavior provided in Niziolek and Guenther (2013) were performed exclusively on the group level, it remains unclear whether a phonemic category change of the target vowel influences the magnitude of compensatory response equally across all speakers.

Our experiment was designed to provide information about how different articulatory and auditory factors, known to have a major influence on the general compensatory response to auditory perturbation on its own, may impact speakers' individual compensatory performance. Thus, we might observe distinct compensatory patterns within and across speakers associated with specific perturbation conditions. In our study, we pursued two goals. First, we wanted to know whether speakers would use two distinct compensatory strategies to produce one vowel, which is characterized by a high degree of linguopalatal contact. More specifically with regard to this question, it is unknown whether speakers predominantly rely on auditory feedback even if it requires them to adopt compensatory strategies which considerably deviate from their usual coarticulatory pattern. Second, we were interested in the question of whether we would be able to observe any systematic individual differences dependent on the phonemic contrast between the produced and the perturbed acoustic signal.

II. METHODS

A. Participants

Thirty-two native speakers of Russian (25 females, 7 males) without reported speech, language, or hearing disorders participated in the study. The participants were recruited from the pool of Russian exchange students and young professionals living in Berlin. The mean age of the group was 25.3 years and participants have spent on average 2.9 years in Germany at the time of the recordings. The study was approved by the local ethics committee and all speakers gave their written consent to participate in the study.

B. Equipment

Each experimental session was recorded in a sound attenuated booth. Participants were comfortably seated in a chair in front of a 19 in. liquid crystal display (LCD) flat screen computer monitor, which served to display the stimuli and experimental instructions. All texts were presented in Russian using Cyrillic font. Participants' speech signal was recorded with a Beyerdynamic Opus-54 neck-worn microphone, perturbed in real-time, and fed back via foam tipped E-A-RTONE 3 A insert earphones, which attenuated the air-conducted sound by approximately 25–30 dB while the microphone gain was set in such a manner that it resulted in an approximate feedback level of 75 dB sound pressure level (SPL). This volume was chosen impressionistically during pilot recordings to strike a balance between listening comfort and masking effect of the bone conduction (cf. overview on bone conduction in Rahman and Shimamura, 2013). Throughout the recordings, the microphone gain was fixed across all participants. In order to shift the *F*2 frequency produced by the participants, tracking and real-time formant perturbation was accomplished with AUDAPTER, which is a C++ real-time signal processing application compiled to a MEX-file executable within a MATLAB environment (cf. for technical details Cai *et al.*, 2010). The correctness of the formant perturbation delivered by AUDAPTER was investigated with the help of an independent MATLAB script, which calculated the formant values of the original and perturbed signals by means of linear predictive coding (LPC) analysis of the signals' cepstra. Subsequently, the results of these calculations were visually inspected at random for all participants (Fig. 1). Based on this procedure, we concluded that despite the applied *F*2 perturbation all formants (*F*1, *F*2, and *F*3) remained present in the modified signal as distinct peaks. Furthermore, we assured ourselves that the *F*2 peak was not interchanged with the *F*3 peak as result of the upward *F*2 perturbation.

Direct measurements were performed to determine the delay of the feedback loop of the perturbation system by comparing the onsets of the acoustic response on the input and the output channels. The average delay amounted to 24 ms [standard deviation (*SD*) = 4 ms]. The original and perturbed signals were digitized and saved with a sampling rate of 16 kHz. Along with audio recordings, AUDAPTER stored data files containing the formant values (*F1*, *F*2, and *F*3) tracked during each stimulus production.

C. Experimental procedure and speech stimuli

For our study we chose Russian, since its vowel inventory includes the high central vowel /ɨ/, which is enclosed within the *F*2 space on each side by the two vowels /i/ and /u/. This constellation allowed us to investigate a two-sided compensation in /ɨ/ with bidirectional perturbation of the *F*2

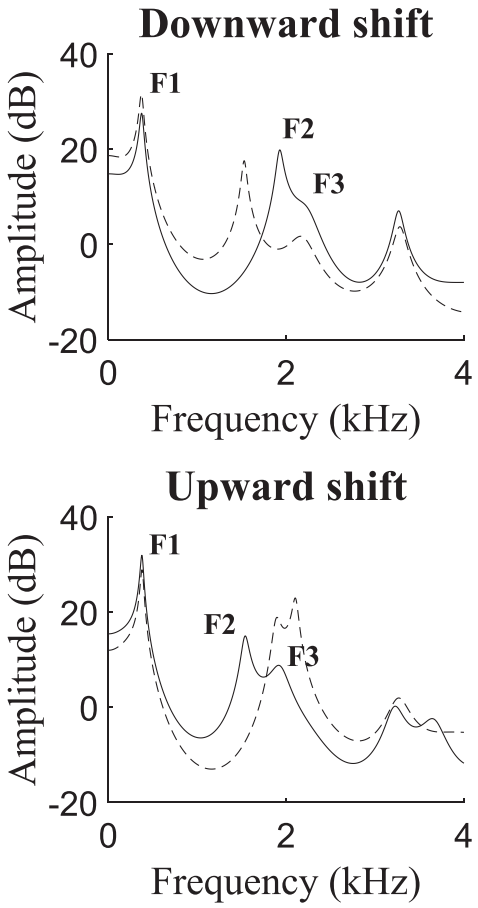

FIG. 1. Example LPC-spectra of the original (solid lines) and perturbed (dashed lines) vowel /ɨ/ during the last shift phase of the experiment.

frequency. The vowel /ɨ/ has a special status in the Russian vowel system since it never appears after palatalized consonants (cf. Bolla, 1981). This detail will become important during the discussion of participants' compensatory behavior in Sec. IV.

Each recording session lasted for approximately 20–25 min and consisted of four experimental phases: baseline, shift 1, shift 2, and shift 3 phases. Before the start of the first experimental phase, participants completed a few practice trials with unrelated speech material to ensure they understood the task and were able to perform it accurately.

During the 60 baseline trials, no auditory perturbation was applied, and on each trial that had an approximate duration of 2 s, participants were visually prompted to produce one of the four CV syllables /di/, /dɨ/, /gɨ/, and /gu/. This was done to assess participants' initial formant space whose structure was expected to have an influence on their compensatory behavior. The interstimulus interval between trials was approximately 1.5 s long. The visual presentation of the stimuli was controlled by a customized MATLAB software package developed at the Institute of Phonetics and Speech Processing, LMU Munich.

During the three following shift phases, of which each lasted for 50 trials, on each trial participants were prompted to produce the high central unrounded vowel /ɨ/ either in alveolar or velar consonantal context (i.e., /dɨ/ or /gɨ/). The *F*2 was shifted during the production of the vowel either upwards or downwards depending on the context (Table I). Within each shift phase, all stimuli were presented in pseudorandom order. That means that a participant could experience one perturbation direction on one trial and the other direction on the immediately following one. On the other hand, the same perturbation direction was never applied on more than two consecutive trials.

TABLE I. Summary of the experimental conditions.

	Group A (Compatible)		**Group B** (Incompatible)	
Constriction	Alveolar (/dɨ/)	Velar (/gɨ/)	Alveolar (/dɨ/)	Velar (/gɨ/)
*F*2 perturbation	Downward	Upward	Upward	Downward

The magnitude of the applied *F*2 perturbation amounted to 220 Hz in the first shift phase and increased incrementally for each shift phase by 150 Hz reaching 520 Hz in the last shift phase. The amount of perturbation did not change within each shift phase. This perturbation scale was chosen on grounds of previous piloting as striking an optimal balance between moderate initial *F*2 perturbation and the experimental goal to let participants learn two strongly distinct compensatory strategies for the vowel /ɨ/.

Participants were instructed to produce all syllables with prolonged vowels. The prolongation of the vowel segments made for one thing the formant tracking more reliable and for the other maximized the amount of time during which participants were exposed to perturbed vowels. To keep the prolongation duration somewhat consistent across participants, they were assisted by a visual go-and-stop signal during their production. The resulting average duration of the produced vowels was 952 ms ($SD = 270$ ms).

Following the experimental session, all participants were asked if they noticed anything unusual in their auditory feedback during the experiment. A few of the participants reported that their pronunciation was different from what they were used to or that they perceived an acoustic difference between the syllables /dɨ/ and /gɨ/. Most participants attributed these pronunciation differences to the effect of listening to own speech on audio recordings, so when asked if and how these differences affected their production, participants reported to have ignored these. From previous research, however, it is known that participants are not able to voluntarily control their reaction to auditory perturbation even if they are told to ignore it (cf. Munhall *et al.*, 2009). Furthermore, participants were asked whether they became aware of any systematic position changes of their articulators (specifically, the tongue position) in the course of the experiment. None of the participants reported to have noticed anything unusual.

D. Interaction between perturbation and coarticulatory effects

The interaction between the place of articulation (alveolar vs velar) and the perturbation direction (upward vs downward) was evenly counterbalanced between the 32 participants, which resulted in two different coarticulatory configurations represented by experimental group A (14 females, 2 males) and group B (11 females, 5 males). Due to coarticulation, baseline

$F2$ was expected to be higher in /dɨ/ compared to /gɨ/. In group A, $F2$ was decreased in /dɨ/ and increased in /gɨ/, such that compensatory movements were expected to act in the same direction as coarticulatory effects [compatible condition; Fig. 2(A)]. In that case, $F2$ values produced for /dɨ/ and /gɨ/ during the baseline phase were expected to drift apart during the shift phases of the experiment due to compensation but remain in the same relation (/dɨ/ > /gɨ/). In group B, the perturbation direction was swapped for both syllables and the baseline $F2$ values were expected to intersect for the two syllables during the shift phases of the experiment [incompatible condition; Fig. 2(B)].

E. Data pre-processing

All recordings of 32 participants amounted to 6720 trials. The onset and offset of the vowel segment produced on each trial were labeled manually based on spectrograms using MATLAB graphic input facilities. Subsequently, the corresponding formant vectors were extracted from the AUDAPTER data files based on the labeled onset and offset boundaries. The middle 50% portion of each formant vector was used to compute the formant means produced on each trial.

F. Baseline formant frequencies

All statistical analyses were performed in R (version 3.4.1; R Core Team, 2017). To understand whether speakers' initial formant space could potentially provide an explanation for the occurrence of certain compensatory patterns, we performed an analysis on non-perturbed trials to derive the mean formant frequencies ($F1$, $F2$, and $F3$) for the vowels contained in the syllables /di/, /dɨ/, /gɨ/, and /gu/. By means of pairwise t-tests, we examined the relation between different formant frequencies of the perturbed vowel /ɨ/ as well as its relations to the neighboring sounds /i/ and /u/. To control for the use of multiple t-tests comparisons, the p-value was adjusted applying the Bonferroni correction. Since for each dependent variable ($F1$–$F3$), three comparisons were made (/di/ vs /dɨ/, /dɨ/ vs /gɨ/, and /gɨ/ vs /gu/), the alpha level of 0.05 was set to $0.05/3 = 0.0167$.

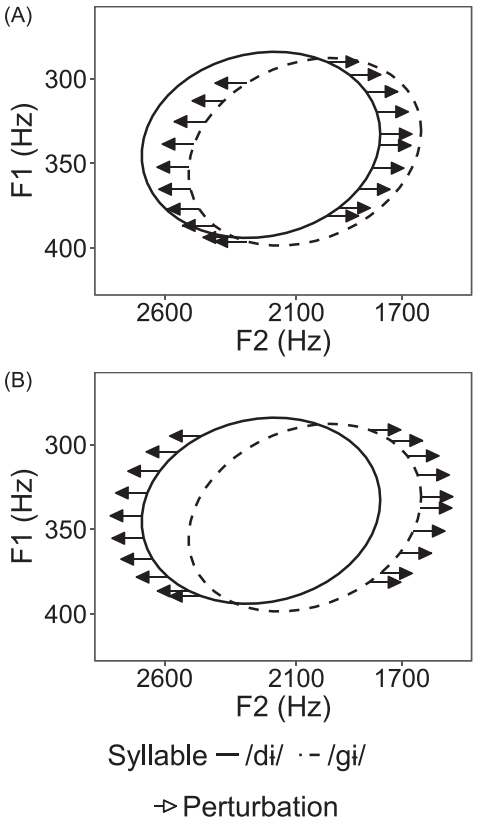

FIG. 2. Distinct perturbation configurations applied for the experimental groups A and B. The ellipses for the two syllables /dɨ/ (solid line) and /gɨ/ (dashed line) are plotted based on the $F1$–$F2$ data of 480 baseline repetitions each. Note that the directions of the figure's axes are reversed. (A) For the two syllables to drift apart, $F2$ was increased in /gɨ/ and decreased in /dɨ/. (B) For the two syllables to intersect, $F2$ was increased in /dɨ/ and decreased in /gɨ/.

G. Adapted formant frequencies

An additional analysis was performed comparing mean formant frequencies ($F1$, $F2$, and $F3$) of the two syllables /dɨ/ and /gɨ/ produced by speakers on non-perturbed trials with formant values produced during the last shift phase of the experiment. By means of pairwise t-tests, we examined how both experimental groups A and B adjusted to the applied perturbation in the target vowel /ɨ/. To control for the use of multiple t-tests comparisons, the p-value was adjusted applying the Bonferroni correction. Since for each dependent variable ($F1$–$F3$), two comparisons were made (/dɨ/ vs /dɨ/ and /gɨ/ vs /gɨ/ for each group A and B), the alpha level of 0.05 was set to $0.05/2 = 0.025$.

H. Analysis of individual compensatory patterns

To identify and classify individual compensatory patterns, we recomputed individual compensation magnitudes reached by participants during the last shift phase of the experiment as percentage scores. This allowed us to subdivide participants into different groups with respect to their compensatory performance for both perturbation directions (upward vs downward).

I. Analysis of general compensatory behavior

To examine average formant changes in participants' production of the two syllables /dɨ/ and /gɨ/ across the four experimental phases, we fitted a generalized additive model (GAM; Hastie and Tibshirani, 1987), which is a significant extension of a generalized linear regression model as it allows the modelling of non-linear relationships between the dependent and independent variables (Wood, 2017a). Therefore, GAMs are much more flexible compared to a linear regression model. The non-linear relationships are modelled via complex functions (smooths) which are constructed from ten basis functions (e.g., linear, quadratic, and cubic functions) with an adjustable number of basis dimensions. The number of basis dimensions is a number which indicates the upper limit of how complex the constructed function can be and is estimated directly from the data during the modelling process. That means that the usage of GAMs does not require a predefined specification of a certain (non-linear) function as it is derived directly from the data. One further advantage of GAMs is the possibility to include random

effects into the model structure to account for individual response variability across but also within speakers. To denote the inclusion of random effects in the fitted model, it is dubbed generalized additive *mixed* model (GAMM). For a hands-on introduction to GAMMs with a focus on dynamic speech analysis, see Wieling (2018).

Prior to building the GAMM model, participants' raw formant frequencies were normalized by subtracting each participant's mean formant frequency produced during the baseline phase for the respective syllable (/dɨ/ or /gɨ/). This was done to exclude participant-specific differences regarding their absolute formants magnitude (e.g., due to gender differences). By means of this normalization, the average *F*1, *F*2, and *F*3 values for /dɨ/ and /gɨ/ were set at zero for the baseline phase.

Subsequently, using the *mgcv* package (Wood, 2017b) we fitted one GAMM model for each formant (*F*1, *F*2, and *F*3) with normalized frequency averaged across all participants and all experimental trials as dependent variable. The data of the unperturbed syllables /di/ and /gu/, which were uttered by participants only during the baseline phase, were not included in the resulting GAMMs. All GAMM models were evaluated, interpreted, and visualized by means of the *itsadug* package by van Rij *et al*. (2017).

In the model structure, we included random factor smooths with an intercept split for the perturbation direction (upward vs downward) in order to assess (potentially nonlinear) individual compensation magnitude differences over the course of the experiment. The model also included a fixed effect which assessed the "constant" effect of the perturbation direction independently from the individual and temporal variation. The interaction between the perturbation direction (upward vs downward) and the experimental group (A vs B) did not significantly improve the model fit, as revealed by the goodness of fit assessed by the Akaike Information Criterion (AIC). Therefore, the data of both experimental groups (A and B) was pooled together for the GAMM analysis.

III. RESULTS

We will start our review of the results by summarizing the initial formant frequencies produced by all participants during the baseline phase of the experiment when no perturbation was applied. After that, we will compare the adapted formant space of experimental groups A and B produced during the last shift phase to the baseline formants. Concluding this comparison, we will discuss different compensatory patterns observed among participants. Finally, we will turn to the presentation of speakers' average compensatory behavior. In particular, we will discuss changes in *F*1, *F*2, and *F*3 frequencies produced by participants over the course of the whole experiment.

A. Initial formant space

The mean *F*1, *F*2, and *F*3 frequencies produced by all participants during the baseline phase are summarized in Fig. 3. Overall, the formants observed in this study for the vowels /i/, /ɨ/, and /u/ were comparable with previous

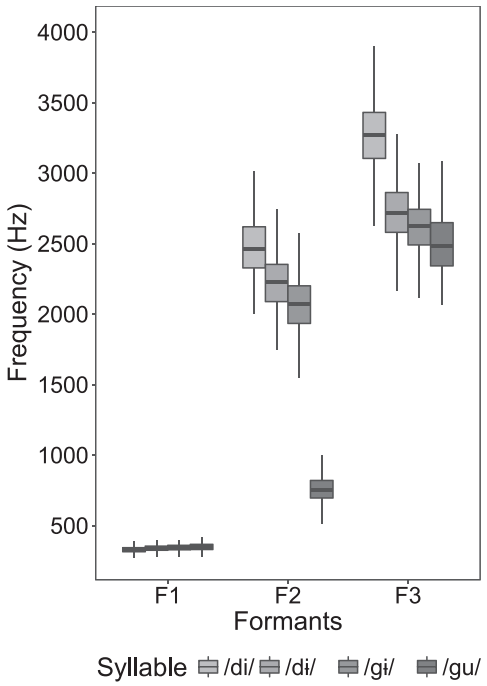

FIG. 3. *F*1, *F*2, and *F*3 frequencies produced by each participant during the baseline phase (no perturbation) for the four syllables /di/, /dɨ/, /gɨ/, and /gu/.

descriptive studies of the Russian vowel space (e.g., Lobanov, 1971). The average within-speaker differences between *F*1 values of the investigated vowels were partially statistically significant, but exhibited rather minor effect sizes, and amounted to −11.84 Hz between /di/ and /dɨ/ [95% confidence interval (CI) (−15.89 −7.80), $t = -5.748$, $p < 0.001$], to −3.80 Hz between /dɨ/ and /gɨ/ [95% CI (−8.06 0.46), $t = -1.750$, $p = 0.08$], and to −4.75 Hz between /gɨ/ and /gu/ [95% CI (−8.77 −0.72), $t = -2.317$, $p = 0.02$]. The *F*2 values, on the other hand, revealed statistically significant within-speaker differences between the investigated vowels which also exhibited prominent effect sizes. The average *F*2 difference between /di/ and /dɨ/ was 241.22 Hz [95% CI (215.67 266.76), $t = 18.532$, $p < 0.001$], 166.92 Hz between /dɨ/ and /gɨ/ [95% CI (140.78 193.07), $t = 12.532$, $p < 0.001$], and 1300.04 Hz between /gɨ/ and /gu/ [95% CI (1278.62 1321.46), $t = 119.140$, $p < 0.001$]. As expected, *F*2 was higher for /dɨ/ compared to /gɨ/ likely due to coarticulation. Regarding the *F*3 values, the three syllables /dɨ/, /gɨ/, and /gu/ were very similar in contrast to /i/, where *F*3 was distinctly higher. Specifically, the *F*3 differences amounted to 556.36 Hz between /di/ and /dɨ/ [95% CI (525.90 586.82), $t = 35.847$, $p < 0.001$], 112.11 Hz between /dɨ/ and /gɨ/ [95% CI (85.27 138.95), $t = 8.196$, $p < 0.001$], and 77.95 Hz between /gɨ/ and /gu/ [95% CI (48.68 107.23), $t = 5.226$, $p < 0.001$].

From Fig. 3, it is apparent that the vowel /ɨ/ is enclosed by its neighboring sounds /i/ and /u/ within the *F*2 dimension. However, there is also an asymmetry with respect to the *F*2 distance between the perturbed vowel /ɨ/ and the upper /i/ on the one hand, and between /ɨ/ and the lower /u/ on the other. Specifically, the *F*2 distance between /i/ and /ɨ/ is lower compared to the *F*2 distance between /ɨ/ and /u/. Furthermore, while the distance between *F*2 and *F*3

frequencies is quite high for both /di/ [−789.53 Hz, 95% CI (−818.84 −760.21), $t = -52.86$, $p < 0.001$] and /gu/ [−1751.29 Hz, 95% CI (−1776.46 −1726.13), $t = -136.63$, $p < 0.001$], the two frequencies are very close together in the perturbed syllables /di̟/ [−474.31 Hz, 95% CI (−501.24 −447.54), $t = -34.671$, $p < 0.001$] and /gi̟/ [−529.20 Hz, 95% CI (−555.34 −503.07), $t = -39.742$, $p < 0.001$]; the implications of these properties of the investigated vowel space on the current findings regarding compensation performance are addressed in the discussion section.

B. Adapted formant space

The mean $F1$, $F2$, and $F3$ frequencies produced by the experimental groups A and B for the vowel /i̟/ during the baseline and the last shift phase are summarized in Figs. 4 and 5, respectively. The average within-speaker changes of $F1$ values were statistically significant for group A and amounted to 11.23 Hz between non-adapted and adapted /di̟/ [95% CI (14.80 7.67), $t = 6.191$, $p < 0.001$] and 11.22 Hz between non-adapted and adapted /gi̟/ [95% CI (15.19 7.25), $t = 5.55$, $p < 0.001$]. For group B, the $F1$ changes were not statistically significant and amounted to 1.59 Hz between non-adapted and adapted /di̟/ [95% CI (8.70 −5.53), $t = 0.438$, $p = 0.66$] and 1.31 Hz between non-adapted and adapted /gi̟/ [95% CI (5.75 −8.36), $t = 0.364$, $p = 0.77$].

For group A, the average $F2$ change between non-adapted and adapted /di̟/ amounted to 17.44 Hz [95% CI (50.52 −14.83), $t = 1.073$, $p = 0.28$] and −216.01 Hz between non-adapted and adapted /gi̟/ [95% CI (−182.93 −249.09), $t = -12.824$, $p < 0.001$]. For group B, the change between non-adapted and adapted /di̟/ amounted to −193.56 Hz [95% CI (−158.97 −228.15), $t = -10.992$, $p < 0.001$] and 20.31 Hz between non-adapted and adapted /gi̟/ [95% CI (57.88 −17.26), $t = 1.062$, $p = 0.28$]. For group A, the $F3$ change between non-adapted and adapted /di̟/ amounted to 43.62 Hz [95% CI (76.09 11.16), $t = 2.639$, $p = 0.008$] and −30.05 Hz between non-adapted and adapted /gi̟/ [95% CI (−0.46 −59.64), $t = -1.995$, $p = 0.04$]. Finally, for group B, the $F3$ change between non-adapted and adapted /di̟/ amounted to −133.74 Hz [95% CI (−95.57 −171.90), $t = -6.885$, $p < 0.001$] and −23.09 Hz between non-adapted and adapted /gi̟/ [95% CI (14.98 −61.17), $t = -1.191$, $p = 0.23$].

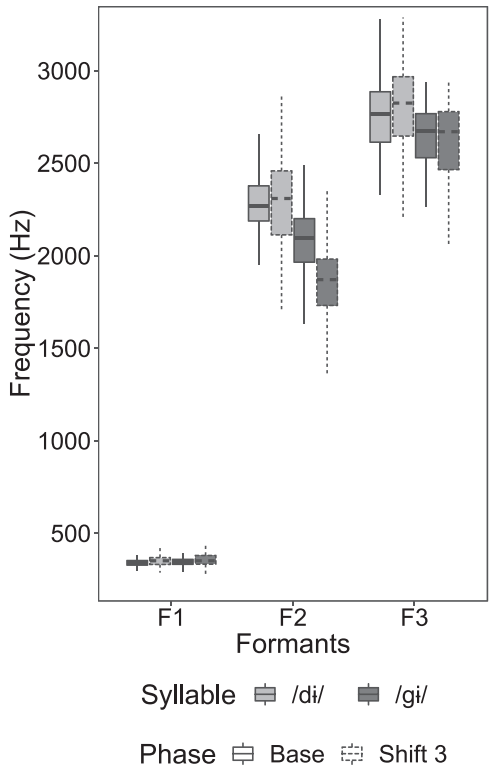

FIG. 4. $F1$, $F2$, and $F3$ frequencies produced by participants of group A during the baseline (no perturbation) and the shift 3 phase (520 Hz perturbation) for the syllables /di̟/ (downward perturbation) and /gi̟/ (upward perturbation).

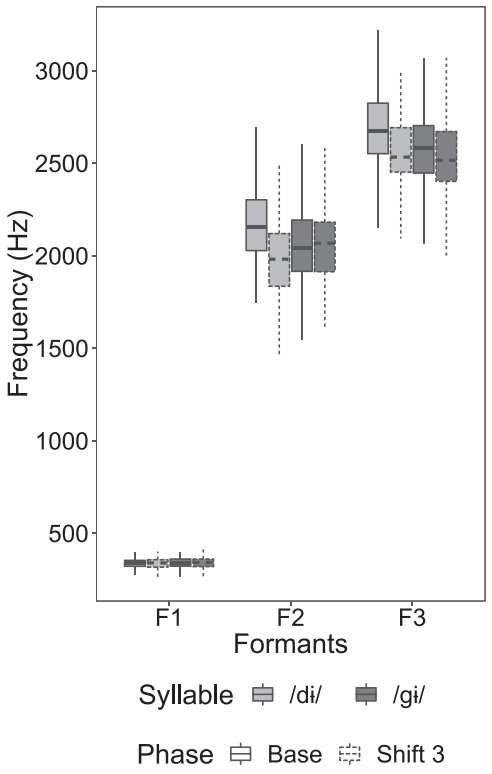

FIG. 5. $F1$, $F2$, and $F3$ frequencies produced by participants of group B during the baseline (no perturbation) and the shift 3 phase (520 Hz perturbation) for the syllables /di̟/ (upward perturbation) and /gi̟/ (downward perturbation).

As predicted, for group A, the initial $F2$ values for the syllables /di̟/ and /gi̟/ drifted further apart by the last shift phase of the experiment such that the $F2$ distance increased on average by 233.45 Hz. On the other hand, for group B, the initial $F2$ values for the syllables /di̟/ and /gi̟/ moved towards each other on average by 173.25 Hz by the end of the experiment. Considering the average $F2$ distance of 124.53 Hz between /di̟/ and /gi̟/ produced by speakers of group B during the baseline phase, we can conclude that, on average, the $F2$ values for both syllables intersected by the end of the experiment by 48.72 Hz.

C. Individual compensatory differences

To get a better overview of individual differences regarding the compensatory magnitude, we recalculated it as individual percentage scores that were reached by participants on average during the last shift phase (Table II). Based on the direction of $F2$ changes as well as the magnitude of

TABLE II. Average *F*2 and *F*3 changes calculated for each participant as percentage scores for the last shift phase of the experiment (520 Hz ≙ 100%). The table includes information regarding participants' coarticulatory configuration (gr.) and the applied perturbation direction (up vs down). Based on their compensatory behavior, speakers were assigned into different groups: the symmetrical (sym.), the asymmetrical (asym.), and the negative (neg.) compensatory pattern (pat.). Three speakers did not display any specific compensatory pattern (-).

		*F*2 (%)		*F*3 (%)					*F*2 (%)		*F*3 (%)		
ID	gr.	*up*	*down*	*up*	*down*	pat.	ID	gr.	*up*	*down*	*up*	*down*	pat.
f1	A	−65	−16	11	−1	asym.	f15	B	−58	6	15	16	asym.
f2	A	5	−11	−14	0	-	f16	B	−3	23	−15	12	-
f3	A	−22	44	−2	27	sym.	f17	B	−40	16	−25	7	sym.
f4	A	−68	26	−37	20	sym.	f18	B	−63	2	−59	−7	asym.
f5	A	−66	9	5	14	asym.	f19	B	−11	24	3	6	sym.
f6	A	−36	4	0	40	asym.	f20	B	−66	−1	13	13	asym.
f7	A	−58	12	−3	3	sym.	f21	B	−54	−9	−76	−28	asym.
f8	A	−90	−67	−7	−17	neg.	f22	B	−37	24	−15	34	sym.
f9	A	−103	6	−8	1	asym.	f23	B	−4	40	−17	−12	sym.
f10	A	−32	−36	−42	−47	neg.	f24	B	−42	−12	−38	−24	neg.
f11	A	44	20	9	25	-	f25	B	−15	16	−3	−1	sym.
f12	A	−54	−25	−9	−23	neg.	m3	B	−31	−1	−24	−17	asym.
f13	A	−45	8	8	13	asym.	m4	B	−72	−21	−36	−13	asym.
f14	A	−19	39	−4	37	sym.	m5	B	−28	−10	−7	4	neg.
m1	A	−24	41	17	43	sym.	m6	B	−40	−28	−83	−39	neg.
m2	A	−30	0	−23	0	asym.	m7	B	−31	4	−47	−15	asym.

the *F*2 difference between both perturbation directions (up vs down) observed for each participant during the last shift phase, we were able to group all participants. This initial grouping was then confirmed visually by examining the slopes of the compensatory changes (Fig. 6).

With regard to *F2* frequency, ten participants exhibited what we dubbed a "symmetrical" compensatory pattern since these participants adjusted their *F*2 values in the opposite direction to the upward and downward shifts by about the same compensatory magnitude [Fig. 6(A)]. Specifically, symmetrical adapters produced the target vowel /ɨ/ with an average compensatory magnitude of 30% ($SD = 21$) and 28% ($SD = 12$) on trials with upward and downward perturbation, respectively. The symmetrical pattern was observed among participants from the experimental group A (five speakers) as well as group B (five speakers).

On the other hand, another 13 participants exhibited an "asymmetrical" compensatory pattern where they on average compensated by 55% ($SD = 21$) for the upward shifts, but by 0% ($SD = 9$) for the downward shifts. That means that asymmetrical adapters produced /ɨ/ under downward perturbation with approximately the same absolute *F*2 values as in their baseline phase [Fig. 6(B)]. The asymmetrical pattern was observed among participants from the experimental group A (six speakers) as well as group B (seven speakers).

In contrast to asymmetrical adapters, another six participants considerably lowered their *F*2 frequency for both of the applied perturbation directions [Fig. 6(C)]. For that reason, we dubbed their compensatory pattern as the "negative" one. These speakers adjusted their *F*2 frequency by −48% ($SD = 23$) during the upward and by −30% ($SD = 21$) during the downward perturbation. The negative pattern was observed among participants from the experimental group A (three speakers) as well as group B (three speakers).

Finally, there were three non-adapters—two speakers from group A and one speaker from group B—who did not exhibit a distinct compensation behavior which would match any of the patterns described above. Furthermore, their reaction to the applied perturbation appeared to be rather unsystematic, so we do not discuss their data any further.

With regard to *F*3 frequency, symmetrical adapters of *F*2 changed it on average by −8% ($SD = 15$) on trials with upward *F*2 perturbation and by +16% ($SD = 18$) on trials with downward *F*2 perturbation. Asymmetrical adapters adjusted their *F*3 values by −17% ($SD = 30$) on trials with upward *F*2 perturbation and by +1% ($SD = 18$) on trials with downward *F*2 perturbation. Finally, negative adapters changed their *F*3 frequency by −31% ($SD = 30$) during the upward shifts and −24% ($SD = 18$) during the downward shifts.

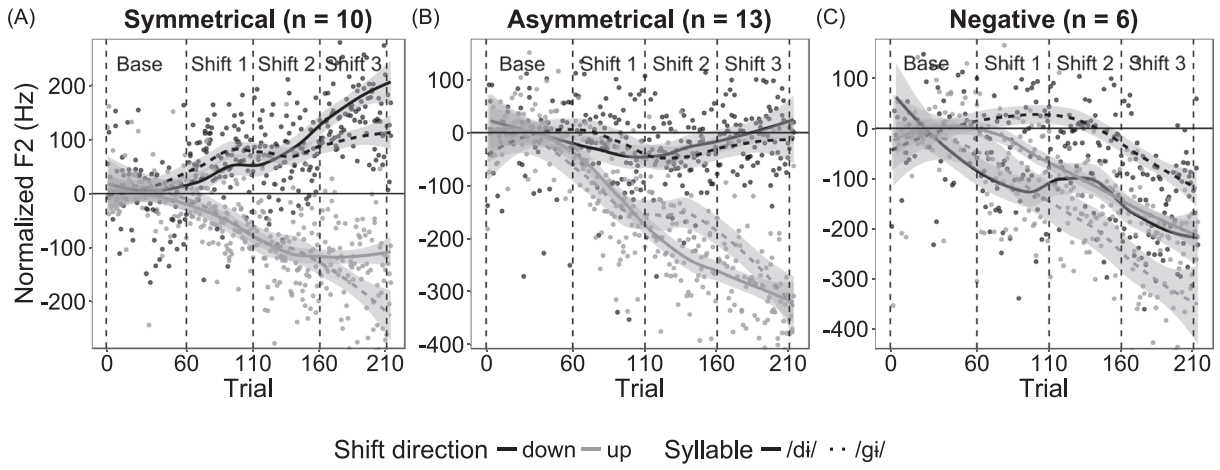

FIG. 6. Average compensatory effects in *F*2 for downward and upward perturbation over the course of the three shift phases. Participants' data is divided into three subplots based on their compensatory pattern: (A) symmetrical adapters, (B) asymmetrical adapters, (C) negative adapters. The data is additionally split by the produced syllable. The plot does not contain the data of the three non-adapters. Individual *y* axis scales were applied due too big differences across compensatory patterns.

D. Relation between *F*2 compensation and *F*3 changes

To understand the relation between the *F*2 compensatory magnitude and the corresponding adjustments of *F*3, we correlated *F*2 and *F*3 changes that occurred during the last shift phase of the experiment (Fig. 7). For trials with upward *F*2 perturbation, Pearson's correlation coefficients revealed a weak, non-significant positive correlation between the compensation magnitudes observed for *F*2 and *F*3 ($r = 0.12$, $p = 0.51$). In contrast, for trials with downward *F*2 perturbation, we observed a strong, significant positive correlation ($r = 0.7$, $p < 0.001$). These findings mean that most participants compensated consistently for the upward *F*2 perturbation by decreasing their *F*2 beyond the corresponding baseline (vertical dashed line in Fig. 7). In this case, no systematic changes occurred to the corresponding *F*3 values which appeared to freely fluctuate around their baseline (horizontal dashed line in Fig. 7). On the other hand, when participants successfully compensated for *F*2 shifts on trials with downward perturbation by increasing their *F*2 values beyond the baseline (vertical dashed line in Fig. 7), it was mostly accompanied by higher *F*3 values.

In Table III, we summarize the Pearson's correlation coefficients for the relation between *F*2 and *F*3 changes independently for the symmetrical, asymmetrical, and negative compensatory patterns. As seen from the table, the three compensatory patterns described in the last section differed systematically with regard to the degree of *F*2–*F*3 correlation observed among them. While symmetrical and asymmetrical patterns were characterized by stronger *F*2–*F*3 correlation on trials with downward perturbation, the participants exhibiting the negative pattern displayed a weaker relation between *F*2 and *F*3 in this case. On the other hand, while asymmetrical and negative patterns were characterized by negative *F*2–*F*3 correlation for trials with upward perturbation, the same *F*2–*F*3 relation held for the symmetrical adapters as on trials with downward perturbation.

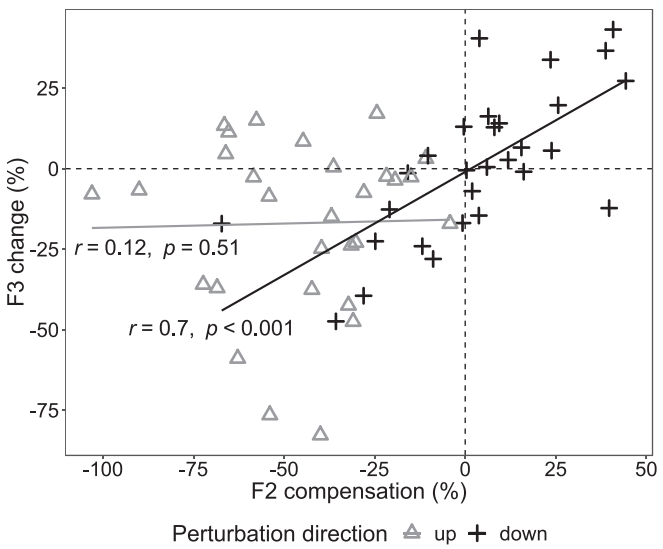

FIG. 7. Correlation between percentage scores of the corresponding *F*2 and *F*3 changes achieved by each speaker during the last shift phase. Correlation is calculated separately for the upward and downward perturbation directions.

TABLE III. Pearson's correlation coefficients for the *F*2–*F*3 relation calculated separately for each compensatory pattern and applied perturbation direction. None of the correlations was statistically significant, likely due to the small number of participants per compensatory pattern and the resulting lack of statistical power.

	*F*2–*F*3 correlation	
Compensatory pattern	*up*	*down*
Symmetrical	$r = 0.51$	$r = 0.46$
Asymmetrical	$r = -0.17$	$r = 0.49$
Negative	$r = -0.39$	$r = 0.22$

E. Average compensatory behavior

1. Effect of the coarticulatory configuration

As previously mentioned in Sec. II G, the interaction between the perturbation direction (upward vs downward) and the experimental group (A vs B) did not significantly improve the fits of the investigated GAMM models. Indeed, the AIC scores were consistently lower for all models which did not contain the interaction between the perturbation direction and the experimental group (−1.2 for the *F*1 model, −4.74 for the *F*2 model, and −1.16 for the *F*3 model). These findings are consistent with the independent analysis of participants' adapted formants presented in Sec. III B and demonstrates further that the information regarding the coarticulatory configuration speakers were assigned to was irrelevant to model their average compensatory behavior. This also likely means that the compensatory behavior did not significantly differ across the experimental groups A and B. To verify this finding, we visually investigated the model fits which included the interaction (figure not given). Since we did not observe any apparent differences in the compensatory behavior of the two experimental groups, below we present the GAMM models which were fitted without the interaction to focus our later discussion on significant findings.

2. F1 changes

The GAMM estimated for *F*1 suggested that the applied perturbation did not have a significant fixed effect on the produced *F*1 values since they did not significantly differ from the baseline either on trials with upward *F*2 perturbation (0.63 Hz, $t = 0.207$, $p = 0.83$) or on trials with downward *F*2 perturbation (3.63 Hz, $t = 1.797$, $p = 0.07$). Taking the temporal variation over the course of the experiment into account, the model did not reveal a *F*1 difference from the baseline for either of the two perturbation directions [Fig. 8(A)]. Random non-linear smooths of the *F*1 model suggest that there were unsystematic participant-specific *F*1 changes which are most likely not related to the applied perturbation [Fig. 8(B)]. Furthermore, a direct comparison between trials with applied upward and downward perturbation revealed no significant difference in their *F*1 curves [Fig. 8(C)]. The average *F*1 difference amounted to −0.61 Hz [95% CI (−6.52 5.31)] by the end of the first shift phase, −0.73 Hz [95% CI (−7.65 6.18)] by the end of the second shift phase, and −0.86 Hz [95% CI (−10.08 8.37)] by the end of the experiment.

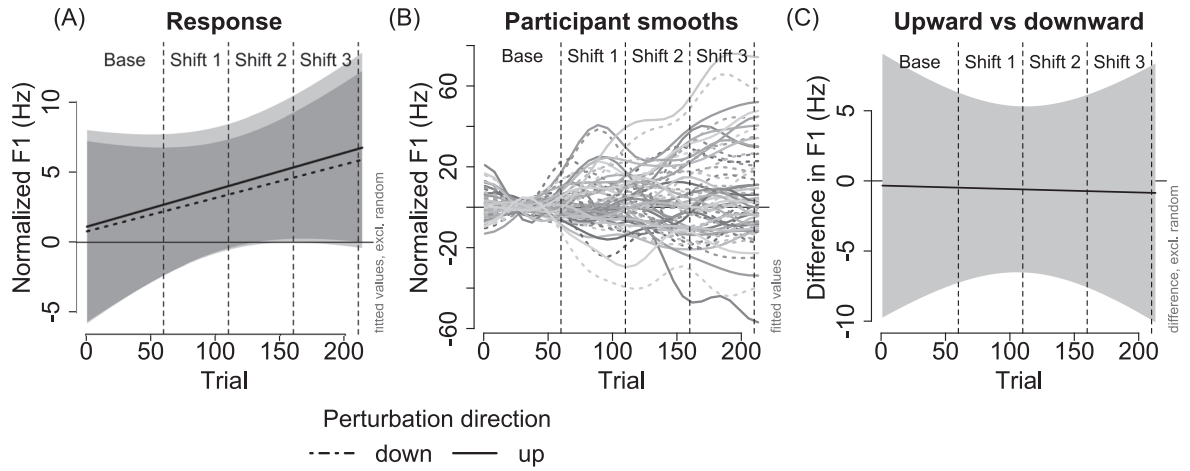

FIG. 8. (A) Average compensatory effects (excluding random participant effects) in $F1$ for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (B) The average difference in $F1$ between the opposing compensatory effects.

3. F2 compensation

The GAMM estimated for $F2$ suggested that the applied perturbation had a significant fixed effect on the produced $F2$ values on trials with upward (−128.06 Hz, $t = -5.616$, $p < 0.001$) but not on trials with downward perturbation (12.12 Hz, $t = 0.961$, $p = 0.33$). However, the direction of the fixed effect was opposed to the direction of the applied perturbation during upward and downward perturbation. Examining the effect of the perturbation over time, the model revealed that the compensation effect increased for both perturbation directions over the course of the experiment [Fig. 9(A)]. Judging from the figure, the effect appears to be on average stronger for the upward perturbation compared to the downward perturbation. The random $F2$ smooths fitted individually for each participant indicate that above and beyond the general tendency to counteract the applied perturbation, participants' compensatory adjustments exhibited high variability in both investigated dimensions [formant frequency and time; Fig. 9(B)]. As indicated by the confidence interval in Fig. 9(C), the $F2$ difference between trials produced under opposite perturbation directions became significant almost immediately at the beginning of the first shift phase and increased, as expected, over the three perturbation phases. The average $F2$ difference amounted to 114.82 Hz [95% CI (62.54 167.10)] by the end of the first shift phase and to 182.64 Hz [95% CI (127.78 237.50)] by the end of the second shift phase. By the end of the experiment, the average $F2$ difference reached 255.74 Hz [95% CI (189.69 321.79)].

4. F3 changes

The GAMM estimated for $F3$ suggested that in the course of the experiment, participants systematically changed their $F3$ values even though only $F2$ perturbation was applied during the experiment. The average fixed effect on the produced $F3$ values on trials with upward $F2$ perturbation was −57.61 Hz ($t = -3.459$, $p < 0.001$) and 7.64 Hz ($t = 0.682$, $p > 0.49$) on trials with downward $F2$ perturbation. The direction of the fixed effect was the opposite of the direction of the applied perturbation during upward and downward perturbation. Examining the effect of the perturbation over time, the model revealed that this effect

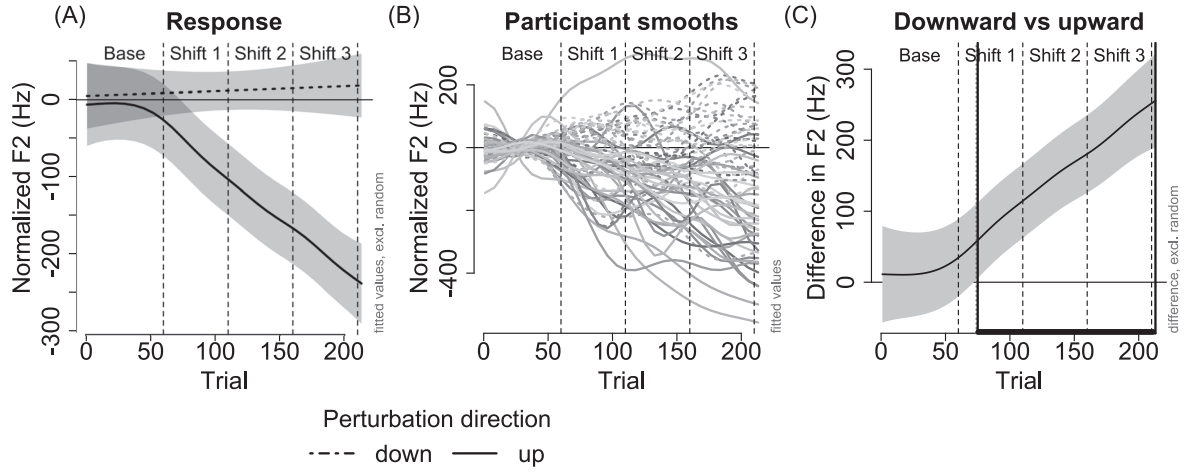

FIG. 9. (A) Average compensatory effects (excluding random participant effects) in $F2$ for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (C) The average difference in $F2$ between the opposing compensatory effects. Solid vertical lines denote the region of significant difference.

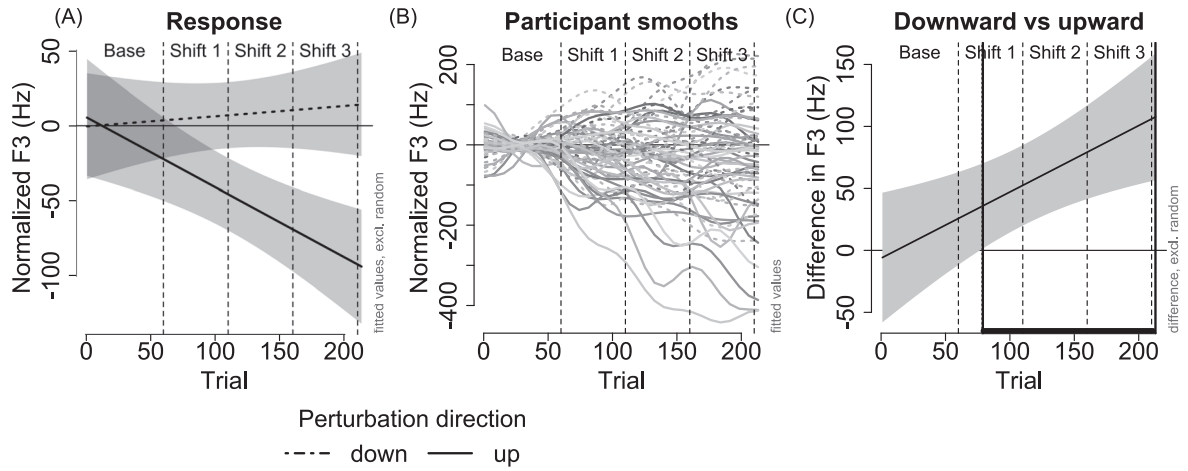

FIG. 10. (A) Average compensatory effects (excluding random participant effects) in *F*3 for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (C) The average difference in *F*3 between the opposing compensatory effects. Solid vertical lines denote the region of significant difference.

increased for both perturbation directions over the course of the experiment [Fig. 10(A)]. The random participant smooths suggest that the magnitude of *F*3 changes varied significantly between participants [Fig. 10(B)]. Similar to *F*2 compensation, it appears that participants' *F*3 adjustments were on average stronger on trials with upward perturbation. As indicated by the confidence interval in Fig. 10(C), the *F*3 difference between trials produced under opposite perturbation directions became significant after the first half of the first shift phase and increased over the three perturbation phases. The average *F*3 difference amounted to 52.56 Hz [95% CI (20.09 82.03)] by the end of the first shift phase and to 80.08 Hz [95% CI (42.01 118.14)] by the end of the second shift phase. By the end of the experiment, the average *F*3 difference reached 107.66 Hz [95% CI (56.59 158.73)].

IV. DISCUSSION

From previous perturbation studies, it is known that speakers can simultaneously use multiple compensatory strategies to achieve the intended acoustic target (Rochet-Capellan and Ostry, 2011; Feng *et al.*, 2011). However, these results are limited to low vowels /ɛ/ and /æ/ and do not generalize since later research established that the magnitude of compensatory responses to auditory perturbation appears to differ across different vowels due to different degrees of physical contact between the tongue and the hard palate involved during the articulation of a particular vowel (Mitsuya *et al.*, 2015). The current study investigated the potential influence of larger linguopalatal contact on speakers' ability to simultaneously use multiple compensatory strategies.

During three shift phases of the experiment, participants produced the high central unrounded vowel /ɨ/, while its *F*2 frequency was perturbed with increasing magnitude of 220, 370, and 520 Hz in opposing directions (upward or downward) depending on the preceding consonant (/d/ or /g/). The bidirectional shift was intended to encourage participants to employ two distinct compensatory strategies to produce the vowel in /dɨ/ and /gɨ/. We investigated two experimental groups, where for the first group *F*2 was decreased in /dɨ/ and increased in /gɨ/, while for the second group, the perturbation directions were swapped for both syllables. This was done to additionally examine the potential influence of the coarticulatory pattern characterizing the relation between syllables /dɨ/ and /gɨ/ on speakers' compensatory production.

Examining participants' average compensatory behavior, we found out that they employed two distinct compensatory strategies depending on the direction of the applied perturbation. This result is consistent with findings made by Rochet-Capellan and Ostry (2011), who demonstrated that participants can employ multiple compensatory strategies for low vowels in the context of *F*1 perturbation. Adding to this result, our data show first that speakers are able to develop multiple compensatory strategies for a vowel with high degree of linguopalatal contact, and second, that speakers adopt multiple compensatory strategies even if these deviate from the coarticulatory relations of their unperturbed speech. These findings demonstrate that auditory feedback serves an important role for online error correction during speech production (cf. Houde and Jordan, 1998; Purcell and Munhall, 2006; Villacorta *et al.*, 2007), and seem to further indicate that speakers disregard somatosensory errors as long as auditory errors are corrected for (cf. Feng *et al.*, 2011). However, a more detailed analysis of individual compensatory strategies among our participants revealed a fine-grained picture hinting at the influence of additional factors on speakers' individual compensatory responses.

In particular, only about 72% of the investigated speakers (23 out of 32) were able to develop two distinct production strategies for the target vowel while the remaining participants failed to do so. Roughly half of the speakers employing two compensatory strategies (10 out of 23) exhibited a symmetrical compensatory pattern, compensating in equal amounts for both applied perturbation directions, and the other half (13 out of 23) exhibited an asymmetrical compensatory pattern, compensating more strongly for the upward perturbation and mostly ignoring the downward perturbation.

At first glance, a hypothesis which could explain the emergence of the asymmetric compensatory pattern might

entail the idea that in this case speakers' compensatory movements were bound by different physical restrictions associated with the two experimental syllables /dɨ/ and /gɨ/. In the case of /dɨ/, for instance, the magnitude of the forward movement of the tongue body, which was required to compensate for downward shifts, was most likely restricted by the alveolar ridge and the upper incisors. On the other hand, the compensation movement in the case of the upshifted /gɨ/ was directed towards the pharynx allowing the tongue to travel a farther distance along the palate. However, as straightforward this hypothesis appears to be, it can account only for a part of the current data since the asymmetrical compensatory pattern also occurred when /gɨ/ was perturbed downwards although there should be no physical restrictions which were comparable to the case of downward perturbation of /dɨ/. In other words, the symmetrical compensatory pattern was observed among speakers independently of whether they were assigned to the experimental group A or B (compatible or incompatible coarticulatory configuration). This observation suggests that the physical restrictions associated with a specific place of articulation of the two syllables /dɨ/ and /gɨ/ (alveolar vs velar) probably did not have a crucial effect on the emergence of symmetrical and asymmetrical compensatory patterns.

Among the 28% of speakers (9 out of 32) who failed to develop two distinct production strategies for the target vowel /ɨ/, six participants exhibited a negative compensatory pattern significantly decreasing their *F*2 frequency irrespective of the perturbation direction, and the remaining three participants failed to compensate consistently for any of the two perturbation directions. However, their production of the vowel /ɨ/ also underwent significant changes in the course of the experiment.

There remains a possibility that changes present in speakers with inconsistent compensatory responses were caused by a formant drift, for instance, due to fatigue. The presence of such formant drift, however, presupposes that a speaker does not actively adjust her/his formants and therefore the observed changes are expected to have rather low effect sizes across all formants (*F*1–*F*3), both produced syllables (/dɨ/ and /gɨ/), and remain independent of the perturbation direction (upward and downward). Contrary to these assumptions, formant changes observed for the three inconsistent adapters were rather heterogeneous with respect to the effect size and the affected formant frequency.

In particular, while there were no significant *F*1 changes in their speech, all three speakers adjusted their *F*2 for at least one perturbation direction by approximately 150–250 Hz, sometimes following the direction of the perturbation. Consequently, we are inclined to believe that these speakers perceived the induced auditory errors but were either not able to classify the directions of the applied *F*2 shifts, and identify the articulatory parameters they needed to adjust in order to restore the intended acoustic goal, or have failed to successfully coordinate the required articulatory adjustments in a manner required for each perturbation direction.

As with symmetrical and asymmetrical compensatory patterns, the negative pattern and inconsistent compensatory responses occurred equally among participants assigned to both experimental groups (compatible and incompatible coarticulatory configuration). This further supports the assumption that physical restrictions associated with a particular syllable in combination with a particular perturbation direction appear to have played but a minor role in the emergence of individual compensatory patterns. This suggests that other factors were responsible for shaping of speakers' individual compensatory responses.

Looking at the individual data ignoring different compensatory patterns, we see that while 90% of participants (29 out of 32) compensated for the upward perturbation, only about 31% of participants (10 out of 32) compensated for the downward perturbation. This compensatory asymmetry seems to be congruent with the asymmetry present in the phonemic space of Russian high vowels.

As described in the introductory section, in Russian, /i/ appears only after palatalized consonants and both /ɨ/ and /u/ follow only non-palatalized ones (cf. Bolla, 1981). Since the dominant perceptual cue of palatalization for Russian speakers is the height of *F*2 frequency at the beginning of the following vowel, it seems reasonable that most participants classified instances of /ɨ/ with increased *F*2 as phonemic errors of palatalization while, on the other hand, less speakers reacted to decreasing *F*2 as it did not induce a change of the phonemic status of the perceived vowel. Additionally, this perceptual effect may have been strengthened by the fact that the *F*2 frequency peak overlapped spatially with the *F*3 peak during upward perturbation making the shifts auditorily more salient compared to downward shifts of *F*2.

Averaging the individual data across all speakers, we arrived at results which appear to be consistent with findings made by Niziolek and Guenther (2013), who showed that speakers react on average much more strongly to perturbations which result in changes of the phonemic category of the perturbed vowel compared to perturbations, and which result only in sub-phonemic changes. Speaking in specific terms, in our case, the compensatory magnitude amounted on average to 45% on trials with upward and to 3% on trials with downward perturbation during the last shift phase of the experiment. However, this is a simplification of the actual compensatory behavior since we know that some speakers (symmetrical adapters) compensated equally for upward and downward perturbation, irrespective of whether it alternated the phonemic category of the perturbed vowel, and other speakers (asymmetrical adapters) reacted essentially only to upward perturbation which alternated the phonemic category of the perturbed vowel.

Considering these detailed observations, it appears that while for symmetrical adapters the perceptual processes involved during compensation related more to vowel discrimination, asymmetrical adaptors relied more strongly on vowel identification (see discussion in Reilly and Dougherty, 2013). In agreement with findings made by Villacorta *et al.* (2007) about the influence of auditory acuity on the compensatory magnitude, we think that symmetrical adapters were presumably more sensitive to *F*2 changes independent of the phonemic status of the perceived vowel.

Another hypothesis which could potentially explain the compensatory asymmetry is that although it was feasible to compensate for the upward *F*2 perturbation in /ɨ/ by lowering exclusively the *F*2 frequency, a compensation for the downward *F*2 perturbation required from speakers to raise their *F*2 along with *F*3 since both frequencies lie quite close in the target vowel /ɨ/. Although speakers should be able to raise *F*2 and *F*3 simultaneously by changing one articulatory parameter, such as the horizontal tongue body position, the adjustment of the *F*3 frequency might be easier to achieve involving additional articulatory changes such as the degree of lip opening or spreading. From previous literature on Russian vowels, it is known that /ɨ/ is normally produced with a slightly wider lip opening compared to /i/, which has much higher *F*3 values (cf. Bolla, 1981). In this scenario, if speakers narrowed their lips additionally to the forward movement of the tongue, that could raise their *F*3 values and contribute to a stronger *F*2 compensation. Some support for this hypothesis is provided by the results of a correlational analysis of *F*2 and *F*3 changes for the two opposite perturbation directions which demonstrated that on the group level, this correlation was significant and highly positive on trials with downward perturbation but not on trials with upward perturbation.

An additional correlational analysis performed separately for speakers exhibiting different compensatory patterns revealed furthermore that the positive F2-F3 correlation held for symmetrical adapters not only on trials with downward, but also on trials with upward perturbation. This suggests that these speakers developed a compensatory strategy which encompassed *F*2 and *F*3 changes for both perturbation directions. On the other hand, asymmetrical adapters developed a compensatory strategy encompassing only the *F*2 frequency, which was sufficient to compensate for the upward perturbation but not for the downward perturbation.

The hypothesis that speakers employ individual compensatory strategies involving different degrees of articulatory complexity may also provide some explanation for the occurrence of the negative compensatory pattern. Judging from the correlational analysis, negative adapters developed a compensatory strategy for the upward perturbation, which was similar to that of the asymmetrical adapters, but used it for both perturbation directions. Although the exact mechanisms behind this remain to be seen, it is possible that negative adapters tried to correct for the perceptually more salient auditory errors, which occurred during the upward perturbation and, at the same time, adhered to some kind of articulatory economy by employing the same compensatory strategy also during the downward perturbation.

The two explanations for the emergence of different compensatory patterns provided from perspectives of speech perception and articulation are not mutually exclusive. On the contrary, following the idea that representations of speech sounds are defined in a multidimensional auditory-somatosensory space (e.g., Hickok, 2012; Sato *et al.*, 2014; Guenther, 2016), it appears plausible that both dimensions could and should have an influence on speakers' ability to compensate for perturbations. Indeed, the presence of different compensatory patterns in our data hints at the idea that speakers might differently weight the information provided by different feedback channels (cf. discussion in Lametti *et al.*, 2012).

In summary, our analyses suggest that although speakers are able to use multiple compensatory strategies even for vowels with larger linguopalatal contact, there is an array of auditory and, probably, articulatory factors that have an additional influence on speakers' individual compensatory performance. While some of the previous work has pointed out the importance of several factors like phonemic relations and non-redundant perceptual cues of the perturbed vowel, the current study provides evidence for the advantage of analyzing individual participants' performances since this may provide deeper insights into the mechanisms of feedback control and the nature of speech targets.

ACKNOWLEDGMENTS

We gratefully acknowledge support by DFG Grant No. 220199 to J.B. We thank the anonymous reviewers for their many insightful comments and suggestions, which considerably improved the manuscript. We thank Felix Golcher for his advice on the statistical modelling of the data. We thank Miriam Oschkinat and Yulia Guseva for their support during data acquisition and acoustic segmentation. We also thank all participants who took part in the study.

Bolla, K. (**1981**). *A Conspectus of Russian Speech Sounds* (Hungarian Academy of Science, Budapest, Hungary).

Bondarko, L. V. (**2005**). "Phonetic and phonological aspects of the opposition of 'soft' and 'hard' consonants in the modern Russian language," Speech Commun. **47**(1), 7–14.

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**(4), 2033–2048.

Feng, Y., Gracco, V. L., and Max, L. (**2011**). "Integration of auditory and somatosensory error signals in the neural control of speech movements," J. Neurophysiol. **106**(2), 667–679.

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

Hastie, T., and Tibshirani, R. (**1987**). "Generalized additive models: Some applications," J. Am. Stat. Assoc. **82**(398), 371–386.

Hickok, G. (**2012**). "Computational neuroanatomy of speech production," Nat. Rev. Neurosci. **13**(2), 135–145.

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

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of *F*0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**(3), 1246–1251.

Katseff, S., Houde, J., and Johnson, K. (**2012**). "Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?," Lang. Speech **55**(2), 295–308.

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**(27), 9351–9358.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**(2B), 606–608.

MacDonald, E. N., Goldberg, R., and Munhall, K. G. (**2010**). "Compensations in response to real-time formant perturbations of different magnitudes," J. Acoust. Soc. Am. **127**(2), 1059–1068.

Mitsuya, T., MacDonald, E. N., Munhall, K. G., and Purcell, D. W. (**2015**). "Formant compensation for auditory feedback with English vowels," J. Acoust. Soc. Am. **138**(1), 413–424.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (**2009**). "Talkers alter vowel production in response to real-time formant

perturbation even when instructed not to compensate," J. Acoust. Soc. Am. **125**(1), 384–390.
Nasir, S. M., and Ostry, D. J. (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**(19), 1918–1923.
Niziolek, C. A., and Guenther, F. H. (**2013**). "Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations," J. Neurosci. **33**(29), 12090–12098 (2013).
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**(2), 966–977.
R Core Team (**2017**). "R: A language and environment for statistical computing (version 3.4.1)," R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (Last viewed July 29, 2019).
Rahman, M. S., and Shimamura, T. (**2013**). "A study on amplitude variation of bone conducted speech compared to air conducted speech," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, October 29–November 1, 2013, Kaohsiung, Taiwan, pp. 1–5.
Reilly, K. J., and Dougherty, K. E. (**2013**). "The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback," J. Acoust. Soc. Am. **134**(2), 1314–1323.
Rochet-Capellan, A., and Ostry, D. J. (**2011**). "Simultaneous acquisition of multiple auditory–motor transformations in speech," J. Neurosci. **31**(7), 2657–2662.
Sato, M., Schwartz, J. L., and Perrier, P. (**2014**). "Phonemic auditory and somatosensory goals in speech production," Lang. Cogn. Neurosci. **29**(1), 41–43.
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**(2), 1103–1113.
Tremblay, S., Shiller, D. M., and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature **423**(6942), 866–869.
van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (**2017**). "itsadug: Interpreting time series and autocorrelated data using GAMMs, R package (version, 2.3)," https://cran.r-project.org/web/packages/itsadug/ (Last viewed July 29, 2019).
Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**(4), 2306–2319.
Wieling, M. (**2018**). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," J. Phon. **70**, 86–116.
Wood, S. N. (**2017a**). *Generalized Additive Models: An Introduction With R* (CRC Press, Boca Raton, FL).
Wood, S. N. (**2017b**). "MGCV: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation, R package (version 1.8-19)," https://cran.r-project.org/web/packages/mgcv/ (Last viewed July 29, 2019).