A functional principal component analysis of acoustic changes under articulatory perturbation

Xinyu Zhang
CLS, Radboud University

The extent and time course of compensation under articulatory perturbation provide a critical test of speech motor control models. This study investigates acoustic adaptation to a bite block using functional principal component analysis (FPCA) of formant trajectories. Fifty-four native Dutch speakers produced pseudowords containing /ɪ/ and /ɛ/ during a pretest and a perturbation phase, with articulatory condition (bite block vs. control) and auditory feedback (normal vs. masking noise) manipulated. FPCA was applied to trial-by-trial trajectories of F1 and F2. The analysis showed no differentiation between bite-block and control conditions in the F1 FPCA space, indicating rapid and near-complete adaptation for vowel height. In contrast, F2 trajectories were affected by the articulatory perturbation: the first principal component separated bite block conditions and revealed gradual, incomplete adaptation across trials. No effects of auditory feedback were observed. These findings suggest dimension-specific compensation, with stable task-level control of F1 and slower adjustment of F2, consistent with models that incorporate both motor equivalence and feedback-based adaptation. FPCA provides a sensitive approach to capturing non-linear dynamics in speech production under perturbation.

1 Introduction

Human speech production relies on the coordinated control of multiple articulators to achieve acoustic goals, but the extent to which this coordination is stable under perturbation remains an open question. Mechanical constraints on the vocal tract such as the insertion of a bite block that restricts tongue movement directly interferes with habitual articulatory patterns, creating a mismatch between intended and achievable movements. Whether and how speakers compensate for such perturbations depends on the underlying control architecture. If speech is governed by stable task-level goals, the system may flexibly recruit alternative articulatory strategies to achieve the motor equivalent of the articulation target and its corresponding acoustic outcome. In contrast, if production relies on learned motor patterns or detailed somatosensory outcome of the vocal tract, perturbations may disrupt both articulation and acoustics, at least

Linguistics in Amsterdam, 2026: 167–185
© Xinyu Zhang

until the system adapts. Bite block paradigms thus offer a critical test case for competing models of speech motor control, as they differentially predict the degree, time course, and nature of articulatory and acoustic change under perturbation.

These theoretical differences are formalized in several prominent models of speech production, which offer different accounts of how the system responds to perturbation. In the Task Dynamics model (Saltzman & Kelso 1987; Saltzman & Munhall 1989), speech is organized in terms of task-level goals defined over vocal tract constrictions, and speech sounds are stored as patterns of gestures and control is expressed in task space. From this perspective, a bite block perturbs the articulators but leaves task goals unchanged, and compensation arises through motor equivalence: alternative articulatory configurations are recruited to achieve the same constriction targets, yielding relatively stable acoustic outputs[1]. According to the model, the dynamics of the articulatory system allow immediate compensation without needing feedback-based error correction, due to the coordinated nature of gestural control. Somewhat similarly, the GEsture shaped by the Physics and by a PErceptually oriented Targets Optimization (GEPPETO) model (Perrier 2014; Patri, Diard, & Perrier 2015) conceptualizes speech production as an optimal control problem over a biomechanical system. Here, speech targets are somatosensory and auditory regions, and execution is largely feedforward. Under the GEPPETO framework, the bite block alters the physical constraints of the vocal tract, requiring the computation of a new motor plan (where the auditory vs somatosensory target weighting is also adjusted due to the articulatory perturbation) that minimizes effort while achieving auditory goals. The model thus predicts multiple compensatory strategies and variations between speakers. In GEPPETO, adaptation may occur but is often incomplete and depends on the weighting difference between somatosensory and auditory targets. More in line with an Optimality Theory (OT) account, Boersma (1998) formulates the relationship between somatosensory and auditory targets somewhat differently. In Boersma's modelling, speech production is (perceptual-) goal oriented, i.e., that the speaker starts with an intended perceptual specification (an auditory goal) instead of aiming for a particular articulatory or somatosensory configuration, and the production emerges from competing constraints rather than from a single fixed motor target. The perceptual goal is not cast in articulatory terms but rather oriented towards the faithfulness of to the specification of said perceptual goal. Under Boersma's framework, adaptive speech behavior is

[1] Regarding "stable acoustic output", Boersma's work (e.g., Boersma 1998) further distinguishes acoustic output from auditory output where they differ on the level of specification. For a straightforward example, see Boersma (1999) for the perception orientation and (different) articulatory implementations of the Dutch <r> and their perceptual result.

changing the implementation, not necessarily changing the goal, indicating the employment of motor equivalence. However, different from the immediate compensation prediction from the above models, either implemented under OT (e.g. Boersma 2011) or neural network (e.g., Boersma et al. 2020), learning and adaptation would occur gradually rather than immediately according to Boersma.

In contrast to models that postulate stored gestural goals, models that emphasize sensory feedback and internal model updating provide yet another perspective. In the Directions Into Velocities of Articulators (DIVA) model (Guenther 1994, 2016), speech motor control is executed via both feedback and feedforward controls. According to the DIVA model, perturbations lead to discrepancies between expected and actual sensory consequences. Such discrepancies are then communicated through sensory feedback, which then drives the error correction. Over repeated trials, the system updates the feedforward commands and adaptation is achieved.[2] Similarly, the State Feedback Control (SFC) model (Houde & Nagarajan 2011) also operates through feedback control. According to the SFC account, speakers continuously estimate the state of the vocal tract by integrating sensory feedback with internal predictions by the state estimator. With the presence of articulatory perturbation, the system would detect the mismatch between predicted and actual somatosensory feedback. The state estimator can then use this mismatch to rapidly update its belief about the current articulator positions and adjust the next iteration of the motor commands. Extending these approaches, the ACT model (Kröger & Birkholz 2007) conceptualizes speech production as the interaction between neural representations and sensorimotor mappings, in which perturbations engage both auditory and somatosensory feedback loops to recalibrate motor plans over time. The Feedback Aware Control of Tasks in Speech model (FACT) model (Parrell, Ramanarayanan, Nagarajan, & Houde 2019), further emphasizes the integration of task-level control with state estimation, proposing that articulatory goals are achieved through a feedback-based control architecture that combines internal predictions with sensory feedback. Within this framework, perturbations such as a bite block would lead to continuous updating of both task execution and state estimation, resulting in adaptive changes in articulatory and acoustic output.

Critically, these accounts differ not only in how compensation occurs, but also in how it unfolds over time. In models that do not rely on iterative feedback-based updating, compensation (and sometimes even adaptation) is predicted to occur immediately. In contrast, models that update motor commands based on error signals derived from iterative somatosensory and auditory feedback predict

[2] However, Brunner & Hoole (2012) seem to suggest a less important role of somatosensory and auditory feedback once the mappings have been set up and can be used in feedforward control.

a more gradual process of adaptation. These differences further imply that model predictions depend on the availability of auditory feedback. The present study therefore examines the trajectory of acoustic change across trials under bite block perturbation, comparing it to productions in an unperturbed condition, while explicitly manipulating access to auditory feedback.

2 Methods

2.1 Experimental design

The data analyzed in the present study were collected as part of a larger experiment investigating production and perception changes during and after articulatory and auditory perturbations (Zhang et al. 2023; Zhang et al. 2026). The research question of interest in the current study focuses on a subset of the aforementioned experiment, specifically the production data collected during the perturbation phase. The goal is to examine whether the trajectory of formant change across the test block differs across experimental conditions. Only the aspects of the experimental design relevant to these trials are described below. The overall flow of the experiment is shown in Figure 1.

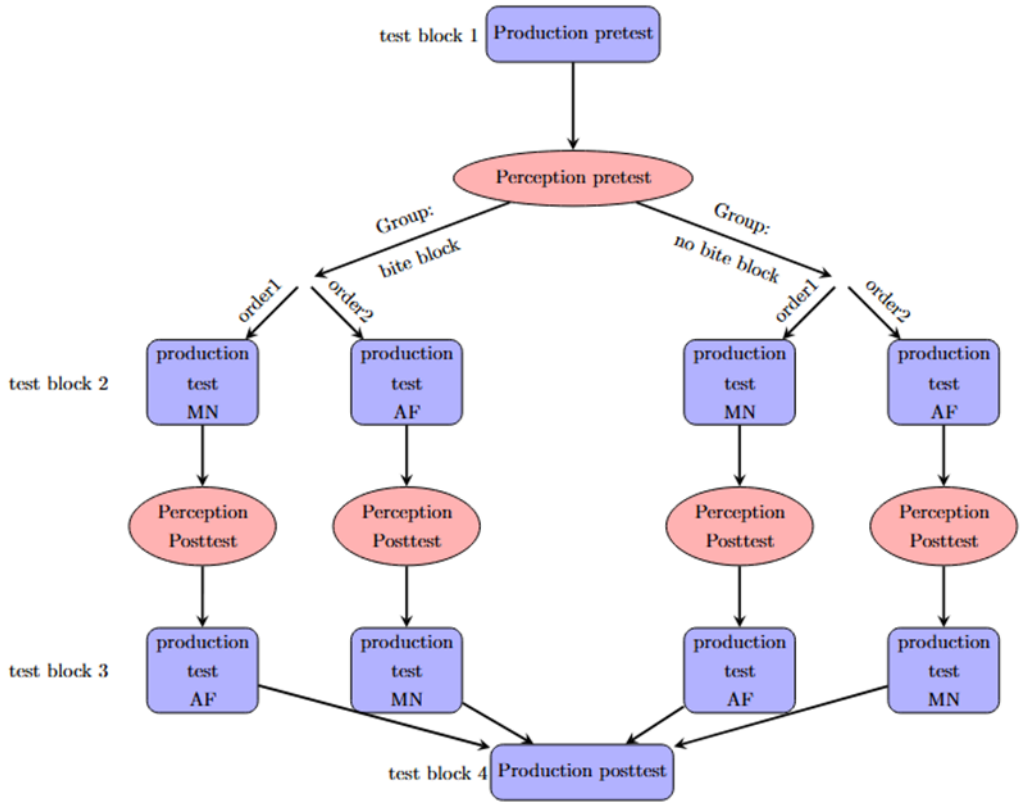


Figure 1: Conditions and orders of the experimental design (pink ellipses represent perception tasks; purple rectangles represent production tasks; MN: masking noise as auditory feedback; AF: regular auditory feedback).

The portion of the experiment analyzed here is the first production test block (labeled test block 1 and test block 2 in Figure 1). During the pretest, all participants produced the stimuli without a bite block and with normal auditory feedback. After the pretest, participants were assigned to different articulatory and auditory feedback conditions.

Participants were first randomly assigned to either a bite block group or a control group (no bite block). Within each group, participants were further randomly assigned to one of two auditory feedback conditions: regular auditory feedback or masking noise.

2.2 Participants

Fifty-four participants between the age of 18 and 30 were recruited through the Radboud Research Participation System. All participants were native speakers of Dutch with normal or corrected-to-normal vision and reported normal hearing and speech abilities (none reported stuttering, other speech disorders, or reading disorders such as dyslexia). Participants received either course credit or a voucher as compensation.

Among all participants, 28 were randomly assigned to the bite block group, 26 to the no bite block (i.e., control) group. Within each of the two groups, between test blocks, participants were randomly assigned to two different orders of auditory feedback conditions (i.e., one test block with regular auditory feedback followed by one with masking noise, or vice versa). Within the bite block group, 15 participants were randomly assigned to the order of regular auditory feedback followed by masking noise (RAF-MN) condition, and 13 were assigned to the reverse auditory feedback order (masking noise followed by regular auditory feedback, MN-RAF). In the group that did not speak with the bite block, 13 were randomly assigned to the RAF-MN condition, and another 13 were randomly assigned to the MN-RAF condition.

Gender was not recorded at recruitment. For the purpose of acoustic analysis (the setting for formant ceilings), perceived vocal gender was coded binarily by a listener (the author) separately after all recordings had been finished. Forty-five participants were labelled as perceived female, and nine were labelled as perceived male.

2.3 Materials

2.3.1 Production stimuli

The target phonemic contrast investigated in the present study was the Dutch /ɪ/–/ɛ/ contrast. This contrast was selected because the primary articulatory difference between /ɪ/ and /ɛ/ involves tongue height, which can be readily affected by a bite block.

The production stimuli consisted of pseudowords conforming to Dutch phonotactics. In each production block, participants produced the same set of pseudowords presented as orthographic prompts on the screen. Given the orthographic transparency of the Dutch language, participants could infer the pronunciation of the pseudowords unambiguously from their spelling. No misreading happened during the experiment.

To obscure the purpose of the experiment and the target vowel contrast, 12 filler items were included in addition to six target items in each production block. All items were repeated three times, resulting in 54 productions per block (6 targets + 12 fillers) × 3 repetitions.

All pseudowords followed the syllable structure C_1VC_2-/ɪf/ (e.g., (e.g., kuufif /ˈkyfɪf/, tessif /ˈtɛsɪf/), such that the target vowel always occurred in a stressed CVC structure. For the six target items, the places of articulation of C_1 were evenly distributed in the target items: /p/, /t/, and /k/ each appeared twice. Each consonant was paired with both target vowels, resulting in six distinct C_1V combinations (3 × 2). The place of articulation of C_2 matched or closely approximated that of C_1. Care was taken to ensure that no C_1VC_2 sequence corresponded to an existing Dutch morpheme, thereby minimizing lexical frequency effects. The six target stimuli are listed in Table 1.

Table 1: Production stimuli used in the experiment

Place of Articulation of C_1	**Target Vowel**	
	/ɛ/	**/ɪ/**
bilabial	*peffif* /pɛffɪf/	*piffif* /pɪffɪf/
alveolar	*tessif* /tɛssɪf/	*tissif* /tɪssɪf/
velar	*keggif* /kɛxɪf/	*kiggif* /kɪxɪf/

Filler stimuli followed the same syllable structure and contained the consonants /p/, /t/, /k/, /f/, /s/ in the C_1 position and /p/, /t/, /k/, /s/, /l/, and /r/ in the C_2 position, and /a/, /ɑ/, /ɔ/, /o/, /i/, /u/, and /y/ as the vowel in the stressed position (V).

2.3.2 Masking noise

Speech-shaped noise was used as masking noise in the masking-noise auditory feedback condition. The noise was created by filtering white noise using a filter derived from the average long-term spectrum of recordings from 10 male and 10 female adult speakers reading a phonetically balanced text. The resulting speech-shaped noise had a duration of 25 minutes and a sampling frequency of 44100 Hz.

2.3.3 Bite block

The bite block used in the experiment was an unopened plastic bottle of a probiotic drink (height 75 mm; smallest diameter 24 mm; largest diameter 38 mm; containing 65 ml of liquid). Each participant in the bite block condition used a new bottle and was allowed to keep it after the experiment.

Participants were instructed to place the bottle with the narrow end facing the inside of the mouth, positioned between the flattened tongue and the palate, and to hold it in place by biting with their incisors on a small indentation on the bottle. When correctly positioned, approximately 36 mm of the bottle extended into the mouth. Figure 2 illustrates the shape of the bite block.

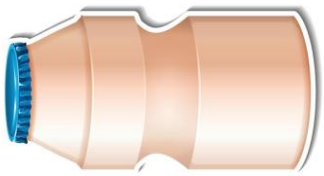


Figure 2: Illustration of the shape of the bite block used in the experiment. Left side of the block was inserted into the mouth. Participants were instructed to bite on the indentation in the middle of the block with their incisors.

2.4 Procedures

The experiment was implemented in OpenSesame. All procedures were conducted in the Center for Language Studies laboratory at Radboud University during a single session lasting less than 20 minutes.

The experiment began with a production pretest, preceded by five practice trials. All production tasks were self-paced: each key press advanced to the next pseudoword prompt. The same set of stimuli was used in all production blocks, with stimulus order randomized separately for each participant. Stimuli were displayed on a BenQ PD2700U monitor (font size 32; resolution 1920 × 1080). Audio recordings were made using a Sennheiser ME 64 microphone, with each token stored as a separate audio file.

Following the production pretest, participants completed a perception pretest, which is not relevant to the present study. Participants then proceeded to the production test block. In this block, participants assigned to the bite block condition produced the stimuli while holding a bite block in their mouth, whereas participants in the control group spoke without it. Within each articulatory condition group, half of the participants had access to regular auditory feedback, and the other half had masking noise during the entire production test block.

In the masking-noise (MN) condition, participants wore Avantree ANC031 headphones with active noise cancellation enabled. The masking noise began

playing after the test block instructions were displayed and before the first stimulus appeared. The masking noise continued throughout the block and stopped at the end of the block.

The masking noise was presented at 92 dBA, measured using a Brüel & Kjær Type 4153 artificial ear. Combined with the active noise cancellation provided by the headphones, this level effectively masked participants' auditory feedback; participants reported being unable to hear their own speech during this condition. In the regular auditory feedback (RAF) condition, participants spoke without headphones and were thus able to hear their own productions.

2.5 Preprocessing

The recordings were automatically annotated using a forced aligner (ten Bosch 2023). The aligner segmented each token into phonemes using a dictionary of labels supplied by the author. The output consisted of Praat TextGrid files containing a stimulus tier and a phoneme tier with temporal boundaries. Because Forced alignment was not able to reliably segment the phonemes of the bite block speech due to the low intelligibility, all annotations for bite block productions were manually corrected by the author.

2.6 Acoustic analysis

Acoustic analysis was done using Praat (Boersma & Weenink 2022). A Praat script [3] read in each audio token together with its corresponding TextGrid, identified the target phoneme (/ɪ/ or /ɛ/ in the first syllable) according to the labels in the TextGrids, and measured the fundamental frequency (F0) as well as the first two formants (F1, F2) of the target vowel. Measurements were taken from the steady-state portion of the vowel, calculated as the 40 - 60% interval of the vowel's duration. Formant objects were calculated using the Burg method, with the maximum number of formants set as 5, formant ceilings set at 5500 Hz for perceived female voices and 5000 Hz for perceived male voices. Time steps were set to automatic, window length was set to 0.025 s, and pre-emphasis was from 50 Hz. Vowel duration was also measured for reference.

2.7 Statistical modeling

Functional principal component analysis (FPCA; Ramsay & Silverman 2002) was used to model the data collected. FPCA is a data-driven method that takes individual curves as input and quantifies each curve in terms of a mean curve and a set of principal component curves (PCs) and their respective weights (PC scores). In essence, each input curve can then be described by a mean curve, plus

[3] Scripts and data are available on the OSF at: https://osf.io/742ec

several ways of deformation from the mean curve (PCs), accompanied by the magnitude of each deformation (PC scores associated with each PC). In the current analysis, each input curve represents the formant variation across the experiment timeline per participant. Namely, each curve captures the formant values of the vowel /ɪ/ as a function of time (trial number) per participant.

FPCA was implemented in R, using the package MFPCA (Happ-Kurz 2025). Data preprocessing was handled by R packages *funData* (Happ-Kurz 2020) and *fda* (Ramsay 2025). Linear regression model fitting was done using the R package *emmeans* (Lenth & Piaskowski 2026).

The input curves were not centered or scaled. For F1, two PCs were retained, since the first two PCs together already account for 100% of the variation in the trajectories. For F2, three PCs were retained, using the criterion of $proportion\ of\ explained\ variance \geq 0.05$. Empty recordings and recordings where clipping happened were excluded, but the original trial numbers were kept (i.e., the location of each token on the trial number axis remained despite of the excluded tokens). In total, 946 input curves were analyzed.

3 Results

3.1 First formant

Figure 3 shows the F1 trajectories of vowel /ɪ/[4] during block 1 and block 2 of the experiment. Color represents bite block condition (BB: bite block, NB: no bite block); Line type represents auditory feedback condition (AF: normal auditory feedback, MN: masking noise as auditory feedback). Test block 1 spanned from trial 1 to trial 54, and test block 2 spanned from trial 55 to trial 107.

[4] Vowel /ɪ/ was plotted instead of /ɛ/ because the tongue height in /ɪ/ is more likely to be affected by the bite block used in the current experiment. However, analyzing vowel /ɛ/ gave similar results.

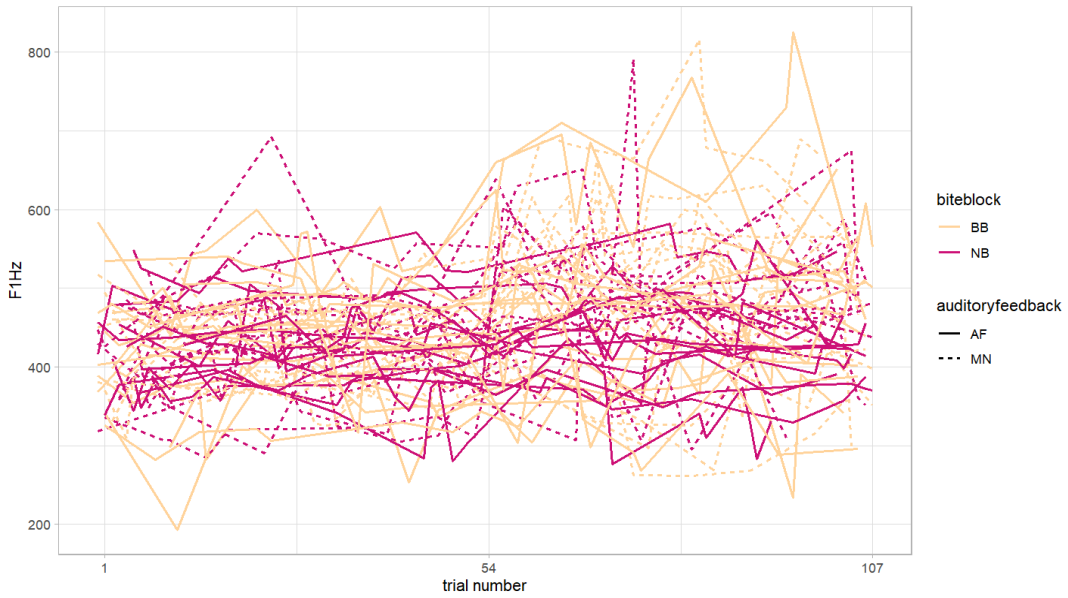


Figure 3: F1 trajectories across trials for all groups, with line colors distinguishing bite block condition (BB: bite block, NB: no bite block), and line types distinguishing auditory feedback type (AF: normal auditory feedback, MN: masking noise as auditory feedback). Trials 1 to 54 belonged to test block 1, and trials 55 to 107 belonged to test block 2.

Figure 4 shows the time-series generated from each of the first two PCs, each plotted across a range of PC scores (proportions of the s.d. of the PC) from -1 to 1. The black dashed line superimposed on each PC series is the mean curve fit to the aggregated data (same curve projected on each PC). These two PCs together account for 100% of the variation in the combined input F1 trajectories: 95.998% for PC1, 4.002% for PC2.

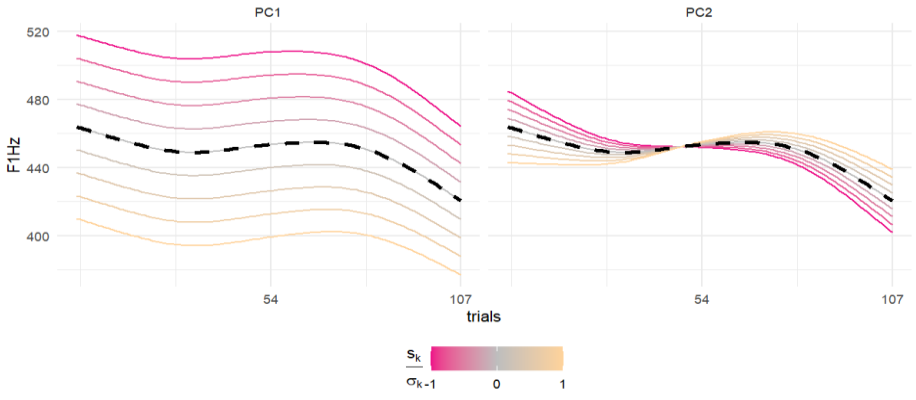


Figure 4: The first two PCs shown as curves that deform the mean F1 trajectory fit to the aggregated data (dashed black line) as determined by the normalized PC score, coded by color. Trials 1 to 54 belonged to test block 1, and trials 55 to 107 belonged to test block 2.

However, when F1 trajectories (i.e., per participant across all trials in the two test blocks) were plotted in the two-dimensional space defined by PC1 score and PC2 score, with color and shape used post hoc to indicate bite block and auditory feedback conditions, neither principal component differentiated between either condition. See Figure 5.

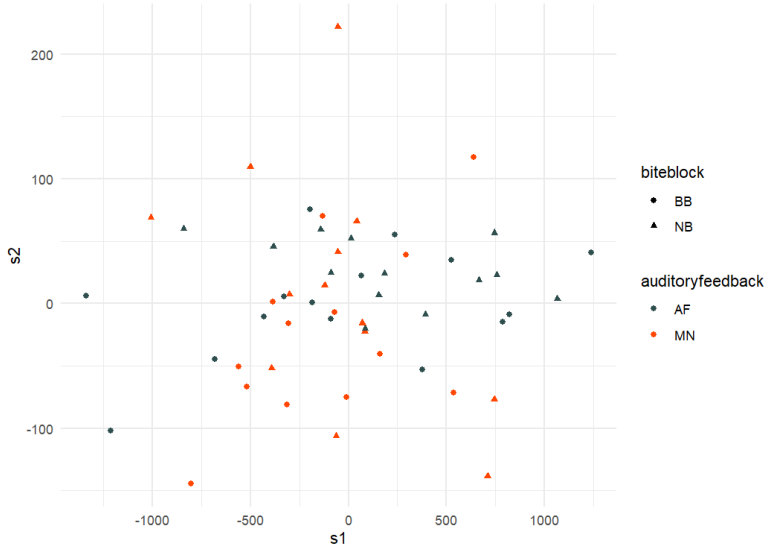


Figure 5: Scatter plot of F1 trajectories according to PC scores (s1: PC score 1, s2: PC score 2). Color- and shape-coding were added post hoc (not provided as input to the FPCA), to indicate bite block conditions. Shape codes bite block condition (BB: bite block; NB: no bite block), and color codes auditory feedback condition (AF: normal auditory feedback; MN: masking noise as auditory feedback).

In summary, the FPCA was not able to capture a difference in the F1 trajectories between the bite block group and the control group, regardless of auditory feedback conditions.

An additional linear regression was run with the PC1 scores as dependent variable, auditory feedback condition, bite block condition, and gender as independent variables. The model used was as follows:

$$lm(s1F1 \sim biteblock * auditoryfeedback + gender)$$

where s1F1 was the score for PC1 in F1 domain. No random intercepts or random slopes were included because the data structure did not include multiple measurements per participant (since multiple F1 measurements of the same participants were linearly arranged into one single curve, thus one participant was one input curve). Factor contrasts for *biteblock*, *auditoryfeedback*, as well as *gender* were sum-coded (-0.5, +0.5).

Results did not show that bite block condition predicted PC1 scores (estimate: 154.37, standard error: 118.76, t = 1.300, p = 0.200), nor did auditory feedback (estimate: -15.44, standard error: 122.00, t = -0.127, p = 0.900). Gender

was shown to strongly predict PC1 scores (estimate: 931.06, standard error: 163.54, t = 5.693, p < 0.001).

The same model with PC2 score as dependent variable showed mild significance in the correlation of bite block condition and PC2 scores (estimate: 30.795. standard error: 17.615, t = 1.748, p = 0.087).

3.2 Second formant

Figure 6 shows the F2 trajectories of vowel /ɪ/ during the first test blocks of the experiment, with bite block condition color-coded, and auditory feedback condition represented by different line types. Both groups spoke without the bite block and with ordinary auditory feedback during test block 1 (trials 1 to 54). In the second test block (trials 55 to 107) the bite block group spoke with the bite block while the control group did not, and within both groups half of the participants had ordinary auditory feedback while the other half had constant masking noise as auditory feedback.

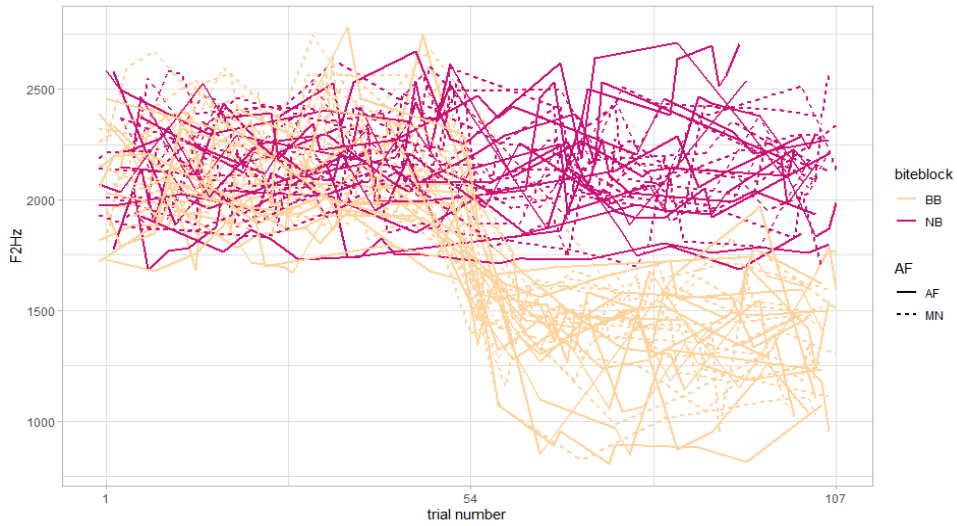


Figure 6: F2 trajectories across trials for all groups, with line colors distinguishing bite block condition (BB: bite block, NB: no bite block), and line types distinguishing auditory feedback type (AF: normal auditory feedback, MN: masking noise as auditory feedback). Trials 1 to 54 belonged to test block 1, and trials 55 to 107 belonged to test block 2.

Figure 7 shows the time-series generated from each of the first three PCs, each plotted across a range of PC scores (proportions of the s.d. of the PC) from -1 to 1. The dashed black line superimposed on each PC series is the mean curve fit to the aggregated data (same curve projected on each PC). Notice that functions with more extreme PC scores (either negative or positive) generated shapes that deform more from the mean curve. These three PCs together accounted for 95.200% of

the variation in the combined input F2 trajectories: 80.864% for PC1, 9.526% for PC2, and 4.810% for PC3.

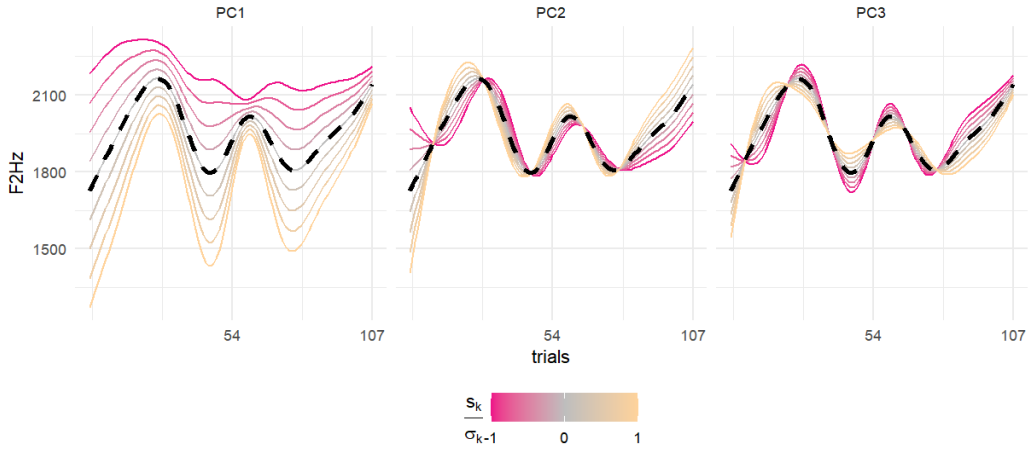


Figure 7: The first three PCs shown as curves that deform the mean F2 trajectory fit to the aggregated data (dashed black line) as determined by the normalized PC score, coded by color. Trials 1 to 54 belonged to test block 1, and trials 55 to 107 belonged to test block 2.

PC1 primarily captured the flatness of the trajectory, with secondary effects on magnitude. Higher PC1 scores rendered the trajectory more turbulent, whereas lower PC1 scores flattened the trajectory. Notice, however, both at the start and end of the second test block, the two types converged more to the mean curve. PC2 primarily captured a phase shift in the curves. Higher PC2 scores resulted in earlier and bigger peaks in F2 variation compared to both the mean curve and the curves from lower PC2 scores. PC3 had a uniform effect of compressing the curves in the time domain with negative PC3 scores and expanding the curves with positive PC3 scores.

Figure 8 plots each participant's F2 trajectory according to the scores of PC1 (s1), the scores of PC2 (s2), and the scores of PC3 (s3)[5]. The colors were added post hoc (not provided as input to the FPCA), to indicate bite block conditions.

[5] Interactive versions of all 3D scatter plots in the paper are available at: https://osf.io/742ec

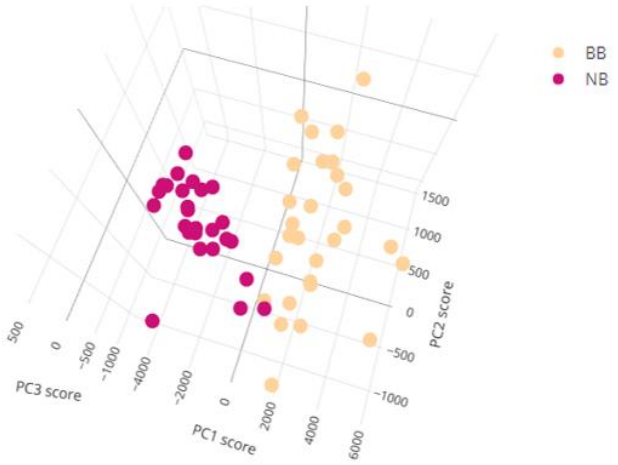


Figure 8: Scatter plot of F2 trajectories according to PC scores. Color-coding was added post hoc (not provided as input to the FPCA), to indicate bite block conditions (BB: bite block; NB: no bite block).

From Figure 8 we can see that the score of PC1 (s1) alone already differentiated the bite block condition, where bite block trajectories tended to have positive s1 and no-bite-block trajectories tended to have negative s1, indicating that the bite block conditions could be differentiated by their F2 values.

In Figure 9, colors were added post hoc to distinguish between different auditory feedback conditions. From the figure, none of the three PCs distinguished the trajectories between different auditory feedback conditions.

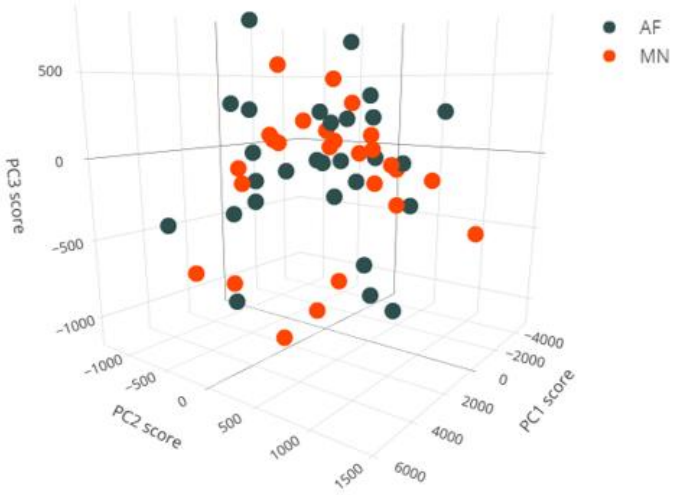


Figure 9: Scatter plot of F2 trajectories according to PC scores. Color-coding was added post hoc (not provided as input to the FPCA), to indicate bite block conditions. Color codes auditory feedback conditions (AF: normal auditory feedback; MN: masking noise as auditory feedback).

As is shown in Figure 8, PC1 was the main factor that differentiated the bite block conditions. Figure 10 shows the typical curves in the two bite block conditions, as predicted by PC1, using the R package *emmeans* (Lenth & Piaskowski 2026).

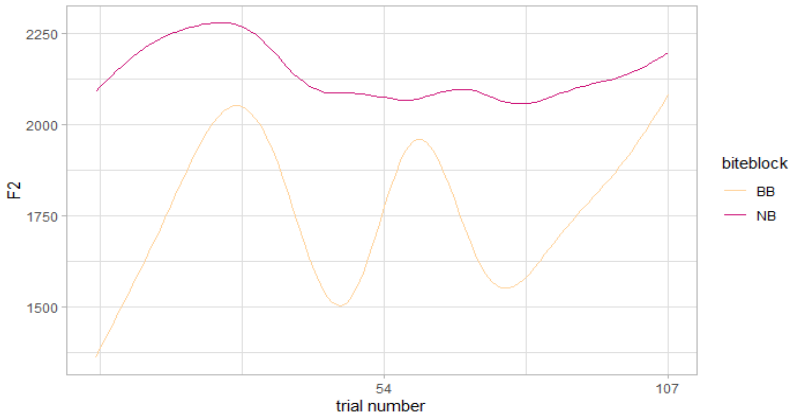


Figure 10: Typical curves of two bite block conditions (BB: bite block, NB: no bite block), predicted by PC1, using *emmeans*. Test block 1 spans from trials 1 to 54, and test block 2 spans from trials 55 to 107.

In Figure 7, trajectories with lower PC1 scores (pink lines in the left panel) show flatter curvature. These trajectories correspond to the group without bite blocks in Figure 8. Similarly, in Figure 10, the no bite block group was predicted to have a more stable trajectory. The bite block group, in contrast, had more variation in F2 across trials. When looking at block 2 (trials 54 to 107) alone, the bite block group started with a lower F2, but by the end of the test block, has adapted to have a higher aggregated F2, closer to the values of the control group as compared to the beginning of the test block.

An additional linear regression was run with the PC1 scores as dependent variable, auditory feedback condition, bite block condition, and gender as independent variables. The model used is as follows:

$$lm(s1F2 \sim biteblock * auditoryfeedback + gender)$$

where s1F2 was the score for PC1 in F2 domain. No random intercepts or random slopes were included because the data structure did not include multiple measurements per participant (since multiple F2 measurements of the same participants were collected into one single curve, thus one participant is one input curve). Factor contrasts for *biteblock*, *auditoryfeedback*, as well as *gender* were sum-coded (-0.5, +0.5).

Results showed that PC1 score was significantly correlated with bite block conditions (estimate: -4152.46, standard error: 301.07, $t = -13.792$, $p < 0.001$) and gender (estimate: 2435.97, standard error: 414.60, $t = 5.876$, $p < 0.001$). No correlation was found between auditory feedback conditions and PC1 score.

Similar models were run with PC2 and PC3 scores as independent variables. For the PC2 model, bite block condition was mildly correlated with PC2 score for F2 (estimate: -341.63, standard error: 152.08, t = -2,246, p = 0.029). No correlation was observed between auditory feedback conditions and PC2 score for F2. In the PC3 model, no correlation between either *auditoryfeedback* or *biteblock* and PC3 scores was found in the results.

In sum, PC1 distinguished the bite block conditions among the input F2 curves. None of the three PCs differentiated between auditory feedback conditions. From the difference in the F2 trajectories between the bite block group and the control group, we can conclude that the bite block group did not achieve full adaptation in F2 under articulatory perturbation. Additionally, the F2 trajectories of the bite block group in combination with the difference between two groups indicate that the compensation under bite block condition was not immediate and was adjusted throughout the entire test phase. FPCA did not capture a difference in trajectories between different auditory feedback conditions for the current data.

4 Discussion

The present study investigated how articulatory perturbation and reduced auditory feedback affect the evolution of F1 and F2 of the Dutch vowels /ɪ/ and /ɛ/ across pretest and test phase of a bite block experiment, with particular emphasis on how speakers adapt over the course of repeated trials. The aim was to examine whether the compensation would happen gradually or immediately, and whether speakers can achieve full adaptation with a tongue height restricting bite block. The results indicate that immediate adaptation is possible for F1, while gradual and incomplete for F2. However, there is no evidence from the current results that auditory feedback makes a difference on the adaptation trajectory. The use of FPCA enabled non-linear modeling of the trajectories, thus revealing results that linear regression alone would not capture.

The immediate adaptation reflected by F1 trajectories aligns with production model accounts in which compensation and adaptation do not rely on iterative feedback-based updating of motor commands. It is thus possible that the somatosensory feedback-based estimation of the current position of articulators, in combination with the coordinated nature of gestural control, is sufficient for typical adult speakers to update the feedforward command and achieve the motor equivalence of desired vocal tract constriction without having to resort to auditory feedback.

The gradual and incomplete adaptation in F2 is more complicated to interpret. It is possible that speakers have a higher weighting for tongue height (the articulatory correlate of F1) than tongue backness (the rough articulatory

correlates of F2) in producing the Dutch vowels /ɪ/ and /ɛ/, resulting in the different extent to which the speakers altered their F1 and F2 during the experiment. Relating this to Boersma's listener-oriented faithfulness account, where production is perception oriented, we should also consider the salience difference between F1 and F2. Given that F1 is perceptually more salient in background noise (Boersma 1998; Bladon & Lindblom 1981) than F2, it is possible, especially under the assumption of the bidirectionality of the Bi-Phon model, that F1 becomes more perceptually specified, thus gains more weight in the production grammar. Formulated under the learning process under constraints reranking, F1 is likely to only have error-driven reranking in one direction whereas for F2 reranking occur in both directions due to transmission noise, leading to the result of the constraints reranking of F1 to finish faster than that of F2.

Additionally, considering the size of the bite block used, and the anatomy of the human oral cavity, it is possible that bite block makes it physically impossible to achieve motor equivalence of an unobstructed /ɪ/ under the perturbation of the bite block. If this physical impossibility is the case, the turbulent trajectory of F2 in the bite block group during the test phase would suggest that the speakers kept searching for alternative configurations of biomechanical constraints (Patri et al. 2015) to determine the optimal motor command to achieve the closest approximate to the motor equivalence (e.g., by raising the larynx). It is also important to note that the bite block itself has an impact on the acoustics measured from the experiment, thus the measured F1 and F2 are not directly the acoustic output of the adaptation process of the tongue but rather that of a combination of the tongue and the bite block. Articulatory data is needed to further clarity.

In terms of the lack of evidence of the role played by auditory feedback, there are two possibilities. First, as mentioned in the discussion of the F1 result, it is possible that adult speakers have already established the articulatory to auditory mapping during the babbling stage of language acquisition, and do not rely as much on auditory feedback in speech production, as Brunner and Hoole (2012) seem to indicate. Second, it is also possible, as predicted by models such as GEPPETO that allow Bayesian updating of internal models and sensory preferences, that speakers under the masking noise condition can rapidly shift the weighting between somatosensory and auditory targets. In this case, speakers might shift the cue weighting once it becomes clear that the masking noise hinders auditory feedback. Additionally, it must be noted that bone conduction might have still been available during the masking noise session of the experiment.

Follow-up studies are needed to investigate the de-adaptation trajectory of such articulatory perturbations.

5 References

Bladon, R. A., & Lindblom, B. (1981). Modeling the judgment of vowel quality differences. *The Journal of the Acoustical Society of America 69*(5): 1414–1422. https://doi.org/10.1121/1.385824

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* [Doctoral dissertation, University of Amsterdam.] https://www.fon.hum.uva.nl/archive/1998/1998-PhD-PaulBoersma.pdf

Boersma, P. (1999). On the need for a separate perception grammar. *Rutgers Optimality Archive*, *358*(1). https://www.fon.hum.uva.nl/paul/papers/percgram.pdf

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In: A. Benz & J. Mattausch (Eds.) *Bidirectional Optimality Theory* (pp. 33–72). https://www.fon.hum.uva.nl/paul/papers/BiPhon21.pdf

Boersma, P., Benders, T., & Seinhorst, K. (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling 8*(1): 103–177. https://doi.org/10.15398/jlm.v8i1.224

Boersma, P., & Weenink, D. (2022). "Praat: doing phonetics by computer". [Computer Program] http://www. praat.org

Brunner, J., & Hoole, P. (2012). Motor equivalent strategies in the production of German/ʃ/under perturbation. *Language and Speech 55*(4): 457–476. https://doi.org/10.1177/0023830911434098

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics 72*(1): 43–53. https://doi.org/10.1007/BF00206237

Guenther, F. H. (2016). *Neural control of speech*. MIT Press.

Happ-Kurz, C. (2020). Object-oriented software for functional data. *Journal of Statistical Software 93*(5): 1–38. doi:10.18637/jss.v093.i05.

Happ-Kurz, C. (2025). *MFPCA: Multivariate functional principal component analysis for data observed on different dimensional domains*. R package version 1.3–11, [Computer Program], https://github.com/ClaraHapp/MFPCA.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience 5*: 82. https://doi.org/10.3389/fnhum.2011.00082

Ramsay, J. (2025). *Fda: Functional Data Analysis*. R package version 6.1.9, [Computer Program], https://github.com/jamesramsay5/fda.

Kröger, B. J., & Birkholz, P. (2007). A gesture-based concept for speech movement control in articulatory speech synthesis. In A. Esposito, M. Faundez-Zanuy, E. Keller, & M. Marinaro (Eds.), *Verbal and nonverbal communication behaviours* (pp. 174–189). Springer. https://doi.org/10.1007/978-3-540-76442-7_16

Lenth, R., Piaskowski, J. (2026). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 2.0.3, [Computer Program], https://rvlenth.github.io/emmeans/.

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS Computational Biology, 15*(9): e1007321. https://doi.org/10.1371/journal.pcbi.1007321

Patri, J.-F., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics 109*(6): 611–626. https://doi.org/10.1007/s00422-015-0664-4

Perrier, P. (2014). "GEPPETO": A target-based model of speech production including optimal planning and physical modeling. *Adventures in Speech Science.*

Ramsay, J. O., & Silverman, B. W.. (2002). *Applied functional data analysis: Methods and case studies*. Springer.

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review 94*(1): 84. https://doi.org/10.1037/0033-295X.94.1.84

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology 1*(4): 333–382. https://doi.org/10.1207/s15326969eco0104_2

Zhang, X., Schoonen, R., & Janse, E. (2023). Sound categorization after speaking with a bite block. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 157–161). ICPhS. https://repository.ubn.ru.nl/bitstream/handle/2066/302324/302324.pdf?sequence=1

Zhang, X., Schoonen, R., & Janse, E. (2026). Effects of somatosensory feedback and auditory feedback on speech under a tongue height obstruction. [Manuscript submitted for publication].

Xinyu Zhang

xinyu.zhang@ru.nl