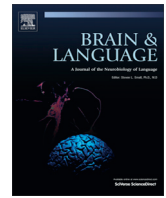




Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Short Communication

Modulations of the auditory M100 in an imitation task

Matthias K. Franken^{*}, Peter Hagoort, Daniel J. Acheson

Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Accepted 4 January 2015

Keywords:

Speech production
MEG
Forward model
Imitation
Individual differences

ABSTRACT

Models of speech production explain event-related suppression of the auditory cortical response as reflecting a comparison between auditory predictions and feedback. The present MEG study was designed to test two predictions from this framework: (1) whether the reduced auditory response varies as a function of the mismatch between prediction and feedback; (2) whether individual variation in this response is predictive of speech-motor adaptation.

Participants alternated between online imitation and listening tasks. In the imitation task, participants began each trial producing the same vowel (/e/) and subsequently listened to and imitated auditorily-presented vowels varying in acoustic distance from /e/.

Results replicated suppression, with a smaller M100 during speaking than listening. Although we did not find unequivocal support for the first prediction, participants with less M100 suppression were better at the imitation task. These results are consistent with the enhancement of M100 serving as an error signal to drive subsequent speech-motor adaptation.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Feedback plays a crucial role in speech motor control as it signals a speaker whether a speech motor action was successful or not. In order to account for an extensive number of findings related to adaptation and feedback processing in motor control, a number of theories have posited a monitoring mechanism that utilizes forward models. Here, motor commands sent to speech articulators also send an 'efference copy' through forward models that predict the somatosensory and/or auditory consequences of those commands (Hickok, 2012; Houde & Nagarajan, 2011; Tian & Poeppel, 2010).

The workings of this internal forward model for speech production are thought to be reflected in a reduction of the auditory cortex response to self-produced speech relative to listening to recordings of the same speech. Magneto- and Electrophysiological studies have found a reduction of the M100, a well-known auditory component that occurs roughly 100 ms after audio onset (Naatanen & Picton, 1987). Theoretical models (Guenther, Ghosh, & Tourville, 2006; Hickok, 2012; Houde & Nagarajan, 2011) explain M100 suppression (M100S) as reflecting a comparison mechanism: if the internal forward model's prediction of the auditory

consequences of speech commands matches the actual auditory input, the cortical response is attenuated. When the prediction does not match the auditory feedback entirely, there is a reduction of M100S (i.e., the auditory cortex shows less suppression). This reduction of M100S then acts as an error signal, driving compensatory mechanisms that adapt motor output towards internal, auditory goals.

Over the last decade, a number of properties of M100S have emerged. Using magnetoencephalography (MEG), Houde and colleagues showed that masking the auditory feedback abolished M100S (Houde, Nagarajan, Sekihara, & Merzenich, 2002). Subsequent studies have shown that the amount of suppression can be modulated by properties of the feedback (Behroozmand & Larson, 2011; Heinks-Maldonado, Nagarajan, & Houde, 2006; Ventura, Nagarajan, & Houde, 2009). Such feedback can also reflect self-produced variation (Sitek et al., 2013). For instance, Niziolek, Nagarajan, and Houde (2013) found that the M100S correlates with the distance between a people's production and the centroid of their vowel space. Together, these studies support a view in which M100S reflects a match between predicted and actual auditory feedback. Additional support for this view comes from direct cortical recordings (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Flinker et al., 2010), where it was also argued that the reduced M100S may be caused either by less neural suppression (i.e., less SIS) or via neural enhancement on top of stable SIS. This neural enhancement was hypothesized to reflect prediction error in the

^{*} Corresponding author at: Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.
E-mail address: m.franken@donders.ru.nl (M.K. Franken).

auditory processing, whereas SIS is helpful in distinguishing self-produced speech from external speech. However, few studies have directly linked M100S to behavioral output (Chang et al., 2013). The current study was designed to more clearly establish this link by having people engage in an imitation task in which auditory feedback is critical to performance, and in which motor consequences of mismatch are likely to adapt in real time.

The goal of this study was to replicate M100S in an imitation task and to test two claims of the aforementioned theories of speech motor control. First, if the M100S indexes the match between a prediction and the incoming auditory signal, then the amount of suppression should relate to the degree of mismatch between the prediction and the auditory signal (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2006; Houde et al., 2002; Liu, Meshman, Behroozmand, & Larson, 2011). Second, if a reduction of M100S serves as an error signal to drive motor adaptation, then individual variation in the amount of suppression should be predictive of the individual variation in imitation aptitude. Specifically, people who show a smaller M100S should show larger adaptations to their speech as they more readily produce error signals that could drive imitation performance.

To test these claims, we measured MEG during online imitation and listening tasks. In the speech imitation task, subjects were instructed to produce the vowel /e/ when a visual cue appeared. At the same moment, subjects heard a recording of themselves producing an auditory stimulus that was the same, close, or far from /e/, and they were asked to imitate this stimulus by adjusting their ongoing vowel production. By varying the acoustic distance between /e/ and the auditory target, we were able to investigate whether this acoustic distance modulated the magnitude of the M100S and people's subsequent speech-motor performance.

2. Results & discussion

2.1. Behavioral performance

In order to assess whether participants appropriately performed the imitation task, an initial analysis focused on people's speech output as a function of time across each imitation condition (see Fig. 1). Results show that for both the first formant (F1) and second formant (F2), participants started at similar values every trial, and the formant values subsequently diverged depending on the vowel. Results of a 9 (Time) \times 5 (Vowel) repeated-measures ANOVA on F1 values (see Fig. 1a) showed significant main effects of both Time ($F(9, 270) = 23.97$; $p < 0.0001$) and Vowel ($F(4, 120) = 145.8$; $p < 0.0001$), as well as a significant Time \times Vowel interaction ($F(36, 1080) = 106.9$; $p < 0.0001$). Similarly, for F2 (Fig. 1b) both main effects as well as the interaction were significant (Time: $F(9, 270) = 9.25$; $p < 0.0001$; Vowel: $F(4, 120) = 79.49$; $p < 0.0001$; interaction: $F(36, 1080) = 60.78$; $p < 0.0001$). With the exception of /ɪ/ condition, all vowels differed from the /e/ condition. Note that although all five vowels were phonemically distinct in Dutch, an important cue to distinguish /ɪ/ from /e/ is vowel duration, a parameter that was lost in this experiment (Adank, van Hout, & Smits, 2004). Importantly, however, these behavioral results demonstrate that participants did imitate the auditory stimuli.

In order to quantify whether the changes in F1 and F2 brought participants closer to the imitation target, we examined the Euclidian distance in F1–F2 space between subjects' speech output and their imitation target on a given trial (Fig. 2c). Results showed significant main effects of Time ($F(9, 270) = 42.1$; $p < 0.0001$) and Vowel ($F(4, 120) = 45.79$; $p < 0.0001$) as well as an interaction ($F(36, 1080) = 34.62$; $p < 0.0001$). Post-hoc tests revealed that for

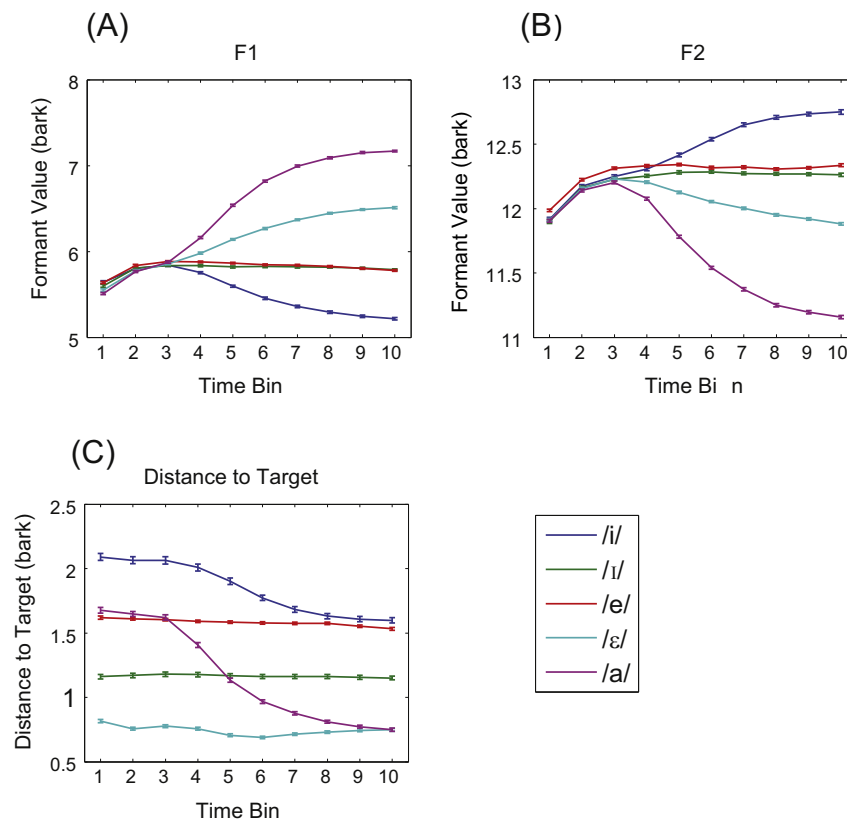


Fig. 1. Average behavioral results across participants. Formant values are expressed in bark (Zwicker, 1961). All error bars represent within-subject standard error of the mean. (A) F1 formant values averaged per time bin and vowel across subjects. Each time bin is 90 ms long. (B) F2 formant values averaged per time bin and condition across subjects. (C) Distance to the imitation target averaged per time bin and condition across subjects.

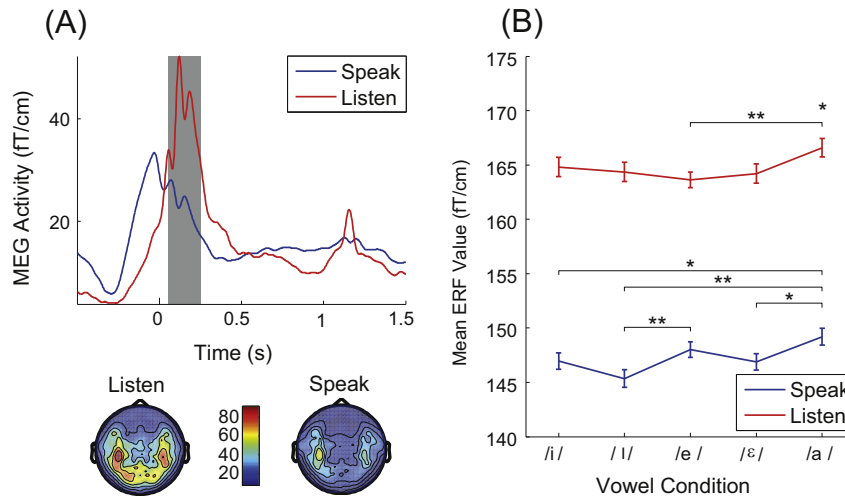


Fig. 2. ERF task effects at the group level. (A) Group-level ERF values per task (all sensors), time-locked to auditory stimulus onset. The time window of interest (50–250 ms) is indicated by the grey shading. Scalp topographies of the listening and speaking ERF (50–250 ms) are shown as well. (B) Group-level ERF averages by Task and Vowel. ERF values were averaged over time (50–250 ms) and channels (subject-specific channel selection, see methods). (*: significant only without multiple error correction; **: significant after Bonferroni correction). Error bars represent within-subject standard error of the mean.

the vowels /i/ and /a/ the distance to the target decreased between time bins 1 and 10 ($t(30) = 6.08$, $p < 0.0001$ and $t(30) = 10.99$, $p < 0.0001$). However, we did not find any effect for the other vowels (all $ps > 0.5$). This may be due to the large individual differences, non-linearities in vowel space, and/or over-adaptations, such as when participants initially get close to the target but then adapt too much, ending farther from the target at the end. Consistent with a previous imitation study (Kent, 1974), these over-adaptations were found for at least one vowel in 14 out of 31 subjects (/a/ – 1 subject; /ɛ/ – 10 subjects; /ɛ/ and /a/ – 3 subjects). No over-adaptations were found for /i/, /l/ or /e/.

2.2. Speaking-induced suppression modulated at the group level

One of the main questions of the present study was whether we could replicate the M100S in an imitation task, and if so, whether the M100S was modulated by the acoustic distance of the stimuli from the starting vowel. Comparison of the ERF in the speaking and the listening task replicated M100S in the imitation paradigm (Fig. 2a). A clear activity peak can be seen in the listening ERF in the time window (50–250 ms) that corresponds to early auditory activity (i.e., the M100), and scalp topographies are consistent with auditory sources. In roughly the same time window, an activity peak is present in the speaking condition with similar topography. Crucially, the listening data clearly shows a larger peak compared to the speaking data. This result thus extends previous work which has observed SIS in simple speech production tasks and in altered feedback paradigms.

It is important to point out that activity was observed prior to stimulus onset (0 ms) in the speaking task (see Fig. 2a). This pre-stimulus activity might reflect an auditory response to speech onset, or possibly speech motor activity, given that stimulus presentation in the speaking condition was triggered by a voice key (i.e., subjects started speaking already before stimulus onset; see methods). Visual inspection of the scalp topography of this early activity (not shown) was very similar to the post-stimulus activity. Consistent with the M100S post-stimulus, this pre-stimulus activity peak was still lower than the post-stimulus listening peak.

In order to examine how the different conditions modulated the ERFs, we calculated the average ERF per task and per vowel. A 2 (Task) \times 5 (Vowel Condition) repeated-measures ANOVA on average ERFs revealed main effects of Task ($F(1, 30) = 85.58$;

$p < 0.0001$) and Vowel Condition ($F(4, 120) = 5.72$; $p < 0.0001$), and a Task \times Vowel Condition interaction that approached significance ($F(4, 120) = 2.05$; $p = 0.09$). In order to investigate the interaction more closely, the amount of suppression was calculated and compared between vowels ($M100S = (ERF_{listen} - ERF_{speak}) / ERF_{listen}$). We found only partial support for our prediction that the magnitude of M100S would scale with deviation from the initial speaking condition. In line with our predictions, the less deviant conditions (i.e. those acoustically closer to /e/), /l/ (mean = 0.11, within-subject SE = 0.0039) and /ɛ/ (mean = 0.10, SE = 0.004), showed numerically more M100S than the more deviant condition, /i/ (mean = 0.11, SE = 0.004) and /a/ (mean = 0.10, SE = 0.004). However, these comparisons were not significant (/l/ vs. /i/: $t(30) = 0.94$, n.s.; /ɛ/ vs. /a/: $t(30) = 0.44$, n.s.). The only comparison that came out significant showed more M100S for /l/ compared to /e/ ($t(30) = 2.88$, $p = 0.007$), which did not correspond to our predictions.

To investigate these comparisons further, the ERF for pairs of vowels were compared separately for each task condition. Fig. 2b shows which comparisons came out as significant (uncorrected; speaking: /i/ vs. /a/, /l/ vs. /e/, /l/ vs. /a/ and /ɛ/ vs. /a/. The contrasts /i/ vs. /l/ and /l/ vs. /ɛ/ were close to significant ($p = .055$ and $p = 0.087$); listening: /a/ vs. every other condition). Corresponding to our expectations, the production ERF in a more deviant condition like /a/ was higher than the ERF for a less deviant condition like /ɛ/, which could correspond to the numerical difference found in the amount of SIS reported above. Note that when the results were corrected for multiple comparisons (Bonferroni), for production only the /l/ vs. /a/ and the /l/ vs. /e/ comparisons survive ($p = .0001$ and $p = .003$). For listening, only the /e/ vs. /a/ comparison survives multiple comparison correction ($p = 0.0004$).

The results in Fig. 2b do not fully correspond to our expectations. One unexpected result is that the ERF for /a/ in the listening condition was larger than the ERF for the other vowels (although mean intensity was equalized, peak amplitude of /a/ was significantly larger than peak amplitude of any other vowel, all $t(31)'s > 2.79$, $p < 0.01$), but maybe even more surprising is the fact that the amount of M100S was higher for /l/ than for /e/. Given that the stimulus in the /e/ condition should be acoustically closest to what participants started with on every trial, this condition should elicit the most suppression and thus the lowest ERF value in the production task. This was not observed, and is difficult to explain.

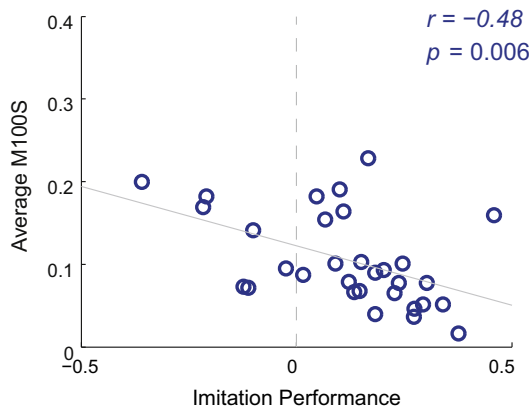


Fig. 3. Correlation between the average imitation performance and the overall M100S per subject. Imitation performance is expressed as the proportional change in distance to the target from the start to the end of a trial ($(\text{distance_to_target}_{\text{end}} - \text{distance_to_target}_{\text{begin}}) / \text{distance_to_target}_{\text{begin}}$). The solid grey line represents a linear fit.

A number of post hoc explanations could perhaps explain this result. First, due to the explicit imitation instructions, people might simply not have expected to hear /e/, because they expected to have to imitate. Second, given that the stimuli were pre-recorded, trial-by-trial variability in the people's initial vowel will have affected the degree of mismatch and therefore the suppression associated with it. Third, if people in our experiment were comparing their motor-based predictions with their actual auditory feedback (rather than the stimuli), neural suppression would be expected not to vary between stimulus conditions. This effect might have been dominant in the M100 modulation, which means the neural suppression would have cancelled out (part of) the error-related neural enhancement. Finally, it is important to note a limitation of our paradigm with respect to this research question. It is possible that during the time window of our analysis, our participants had already issued motor commands to change their ongoing vowel production, so their internal target was not /e/ anymore. However, similar analyses in an earlier (50–150 ms) time window revealed a similar pattern of results. Obviously, these are all merely post hoc explanations and require further investigation.

2.3. SIS predicts participants' imitation performance

Despite the fact that M100S did not scale according to the deviation from people's initial speech production, we can still address whether modulation of the M100S is predictive of how much people adapt their speech motor output. To explore this issue, we investigated whether the average amount of M100S across subjects is correlated with subjects' performance on the imitation task (Fig. 3). We hypothesized that people would vary in their sensitivity to errors, and that this variation would be reflected in the average amount of suppression they show. Less suppression on average should then relate to a higher sensitivity to errors (or being less lenient to prediction errors), which would lead to stronger and/or quicker adjustment.

Corresponding to this prediction, we found that M100S correlated with imitation performance ($r(29) = -0.48$; $p < 0.01$). However, note that participants listened to different stimuli (their own voice), so it would be possible that participants who show less M100S simply heard stimuli that were acoustically further apart. A control analysis showed that the correlation reported above was not due to subject-specific dispersion of the stimuli in vowel space. Specifically, a multiple regression analysis including imitation

performance and stimulus dispersion as separate predictors showed a main effect of imitation performance ($t(28) = -2.99$, $p < 0.01$), but not of stimulus dispersion ($t(28) = 0.58$, n.s.).

Of note in the current results, seven subjects show a mean negative imitation performance. This effect may reflect non-linearities in vowel space, as subjects could have ended up farther away in acoustic space while still staying within the same perceptual category. Importantly, when these subjects were excluded, we still found a negative trend ($r(22) = -0.39$, $p < 0.06$). Also, note that an important limitation of our study is that we are unable to determine whether less M100 suppression actually reflects less neural suppression, rather than an enhancement of neural activity. So although we found a reliable correlation between M100S and behavioral performance, the neural mechanism driving this correlation remains unclear.

3. Conclusion

In the current study we investigated M100 modulation in the context of a speech imitation task. Based on current theories of speech-motor control (Guenther et al., 2006; Hickok, 2012; Houde & Nagarajan, 2011), we predicted that the magnitude of M100S would decrease with the acoustic distance from the starting vowel of people's speech production, and furthermore, that M100S would predict how much people adapted their speech in the imitation task. Acoustic measures of the people's speech output during the task showed that subjects complied with the task instructions by adapting their speech to the auditory stimuli. In the MEG data, we found a strong suppression of the auditory M100 in the speaking task relative to the listening task, which replicates SIS in this new paradigm. A closer look at the ERFs in the different conditions revealed that the varying auditory stimuli modulated the amount of suppression, although the results did not correspond to our predictions.

With regards to the second prediction, an analysis of individual differences in the speech imitation task confirmed a relationship between M100S and participants' performance in the imitation task. Specifically, people with reduced M100S were better at speech imitation. This result is consistent with the models of speech motor control described in the introduction. Here we take the average amount of people's M100S as indicative of their sensitivity to errors. Specifically, some people might be more sensitive (or less lenient) to small errors between their intended speech and acoustic mismatches. As a result, these participants would show on average less M100S. A higher sensitivity to mismatches would lead people to generate error signals more readily, thus driving speech motor adaptation, resulting in better imitation performance.

The findings from the current study are the first to explicitly link individual variation in M100 modulation to speech imitation performance. This result provides an important test of one of the main hypotheses stemming from current theories of speech motor control, and is consistent with M100S indexing an error signal that is used to adapt speech motor behavior. Despite this result, we did not find strong evidence in favor of another prediction stemming from these accounts, namely, that the amount of M100S should decrease with the distance from the original target utterance. As noted above, this may be a result of the current task conditions, and thus remains an open question for future investigation. Nevertheless, the current study highlights the importance of taking individual differences into account when studying speech production and opens up possibilities for further investigations in how inter-individual variation in speech production relates to variation in the psychological and neural mechanisms underlying speech motor control.

4. Methods

4.1. Participants

Thirty-two healthy volunteers (age: $M = 21.8$, $SD = 5.2$; 21 females) participated after providing written informed consent in accordance with the Declaration of Helsinki and the local ethics board committee (CMO region Arnhem/Nijmegen). All subjects had normal hearing, normal or corrected-to-normal vision, were right-handed and were native speakers of Dutch.

4.2. Stimuli

Auditory stimuli included subject-specific vowel recordings made in a preceding session. Participants were recorded while producing the Dutch vowels /i/, /ɪ/, /e/, /ɛ/ and /a/. The vowels were equalized in length (to 1s), in pitch (to the subject's average pitch) and in intensity (to 82 dB). All audio editing was done in Praat (Boersma & Weenink, 2013). Catch stimuli were made by shifting the pitch in the final 100 ms of the sound up by 8 mel. The mel scale is a psychoacoustic scale first proposed by Stevens, Volkman and Newman (1937).

4.3. Experimental design and procedure

Subjects alternated between an imitation task and a listening task in a blocked design. Imitation trials started with a fixation cross (1260–2260 ms, jittered). Participants were instructed to start saying /be/ as soon as three exclamation marks appeared. Participant's speech onset triggered playback (average delay = 51 ms, $sd = 4.3$ ms) of an auditory stimulus (9 x /i/, /ɪ/, /e/, /a/ and 12 x /e/). The first block was a baseline block consisting of 72 /e/ trials. The participant's task was to imitate the auditory stimulus as accurately as possible by adapting his/her pronunciation of the vowel to match the auditory stimulus. The exclamation marks remained visible for 1500 ms after speech onset. The inter-trial interval was 1000 ms.

The listening trials were similar, except the stimulus played directly when the exclamation marks appeared, and participants did not have to vocalize. To maintain focus on the stimuli, participants performed an auditory detection task in which they pressed a button when they detected a rising pitch at the end of the stimulus. Four such trials were added to each block, and feedback for accuracy was provided after each catch trial. The experimental design was implemented using a PC running Presentation® software (Version 16.2, www.neurobs.com). Participants' speech was recorded with a single microphone in the magnetically shielded room with a sampling rate at 44100 Hz and all auditory stimuli were presented binaurally via MEG-compatible air tubes. The complete experiment had 872 trials, which were randomized within blocks.

4.4. Data acquisition

We used an MEG system (VSM/CTF systems, Port Coquitlam, Canada) with 275 axial gradiometers. Three localization coils, fixed to anatomical landmarks (nasion, left and right preauricular points), were used to determine head position relative to the gradiometers. Head position was monitored online by the experimenter and if necessary corrected between the experimental blocks.

All data were low-pass filtered by an anti-aliasing filter (300 Hz cut-off), digitized at 1200 Hz and stored for offline analysis. Participants were seated upright, with the head rested against the back of the helmet and touching the top of the helmet. A small cushion

was used to fix the head's position so as to minimize free head movement.

4.5. Analyses

4.5.1. Behavioral

In the participants' speech recordings, speech onset served as a marker for the beginning of the epoch of interest. The first 900 ms of each epoch were divided into 10 time bins of 90 ms, for which the average F1 and F2 values were calculated using an iterative Burg algorithm. Formant values were expressed in the psychoacoustic bark scale (Zwicker, 1961). The acoustic distance to the stimulus (i.e. the imitation target) was quantified as the Euclidian distance in F1/F2 space between the subject's speech and the stimulus. Two-way repeated-measures ANOVAs with factors Time Bin and Vowel were carried out on the F1 data, the F2 data and the acoustic distance separately. One participant failed to perform the task appropriately and was excluded from all further analyses.

4.5.2. MEG

All MEG data analyses were performed using the Fieldtrip toolbox for EEG/MEG-analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data was resampled at 300 Hz, and epochs of interest were defined from 1s before to 2s after stimulus onset. Trials were excluded if the range and variance of the MEG signal differed by at least an order of magnitude from the other trials of the same subject. Imitation trials in which subjects failed to speak or spoke too softly (therefore not triggering the stimulus playback) were discarded as well. On average, 7 trials per subject were rejected, resulting in an average of 827 trials per subject. The datasets for two subjects lacked one imitation block and one participant's dataset lacked both an imitation and a listening block. As these still contained 792 and 744 trials before trial rejection, we did not exclude their data.

For artifact removal, we performed an independent components analysis (ICA; see Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). Prior to ICA, a principal components analysis (PCA) was run on the data to reduce the number of dimensions to 100. Visual inspection of the ICA components' time courses and scalp topographies, resulted in rejection of an average of 13 components ($SD = 6$) per subject. Components were identified as containing artifacts when they showed clear signatures of eye movement, heartbeat (Vigario, Sarela, Jousmaki, Hamalainen, & Oja, 2000), or speech-related muscle artifacts. The latter type were identified as having peripheral topography (indicative of sources close to the jaw muscles) and a time course that correlated well with speech on- and offset (a step-like response from the moment of speech onset until speech offset). Subsequently, the data was transformed back to sensor space and band-pass filtered using a two-pass Butterworth filter (order: 4; pass-band: 1–12 Hz).

Event-related fields (ERF) were calculated by averaging the trials time-locked to stimulus onset (baseline: 400–200 ms before stimulus onset). All further analyses were done on the planar gradients (Bastiaansen & Knosche, 2000). Statistical testing was done within a time window of interest from 50 ms to 250 ms after stimulus onset. An ANOVA (factors Task and Vowel) was performed on the ERF data averaged across the time window. To select which channels to use, a data-driven approach was performed by selecting the channels that were among the 80 most active channels within the time window in both the listening and the speaking data for each subject (mean = 53; $SD = 7$).

4.5.3. Individual differences

The amount of M100S was calculated per condition as the difference between the ERF values for the listening and speaking trials, normalized by the ERF values in the listening data (i.e.,

(listening–speaking)/listening). The results were averaged across conditions. Imitation performance was quantified as the normalized change in acoustic distance to the stimulus: the distance between the participant's speech and the auditory stimulus was calculated for the first and the last time bin of the subject's speech. Performance was then calculated as the difference between the first time bin and the last bin per condition, normalized by the distance to target in the first time bin (i.e., $(\text{distance_to_target}_{\text{bin1}} - \text{distance_to_target}_{\text{bin10}}) / \text{distance_to_target}_{\text{bin1}}$). Finally, the results were averaged across conditions. So, a value of 0 for behavioral performance indicates the subject are just as far from the target at the end as in the beginning of the trial, whereas a value of 1 indicates perfect imitation. An additional analysis was run to see whether the average amount of suppression was correlated with stimulus dispersion in vowel space. The latter was quantified as the average Euclidian distance to each participant's centroid, the point in vowel space defined by the average F1 and F2 values across all five stimuli.

References

- Adank, P., van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America*, 116(3), 1729–1738.
- Bastiaansen, M. C. M., & Knosche, T. R. (2000). Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clinical Neurophysiology*, 111(7), 1300–1305.
- Behroozmand, R., & Larson, C. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, 12(1), 54.
- Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer Program] (Version 5.3.52).
- Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), 2653–2658.
- Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *Journal of Neuroscience*, 30(49), 16643–16650.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport*, 17(13), 1375–1379.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5(28).
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125–1138.
- Kent, R. D. (1974). Auditory-motor formant tracking: A study of speech imitation. *Journal of Speech and Hearing Research*, 17(2), 203–222.
- Liu, H. J., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, 122(5), 951–957.
- Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, 94(20), 10979–10984.
- Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound – A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, 33(41), 16110–16116.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011(2011).
- Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS ONE*, 8(12), e82925.
- Stevens, S., Volkman, J., & Newman, E. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. *Journal of the Acoustical Society of America*, 8(3), 185.
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, 1, 166.
- Ventura, M., Nagarajan, S., & Houde, J. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10(1), 58.
- Vigario, R., Sarela, J., Jousmaki, V., Hamalainen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *Ieee Transactions on Biomedical Engineering*, 47(5), 589–593.
- Zwicker, E. (1961). Subdivision of audible frequency range into critical bands (Frequenzgruppen). *Journal of the Acoustical Society of America*, 33(2), 248.