Individual variability as a window on production-perception interactions in speech motor control[a]

Matthias K. Franken[b]
Donders Institute for Brain, Cognition and Behaviour, Center for Cognitive Neuroimaging, Radboud University, P.O. Box 9101, Nijmegen, 6500 HB, The Netherlands

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

James M. McQueen and Frank Eisner
Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, P.O. Box 9104, Nijmegen, 6500 HE, The Netherlands

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

(Received 29 February 2016; revised 17 July 2017; accepted 25 September 2017; published online 11 October 2017)

An important part of understanding speech motor control consists of capturing the interaction between speech production and speech perception. This study tests a prediction of theoretical frameworks that have tried to account for these interactions: If speech production targets are specified in auditory terms, individuals with better auditory acuity should have more precise speech targets, evidenced by decreased within-phoneme variability and increased between-phoneme distance. A study was carried out consisting of perception and production tasks in counterbalanced order. Auditory acuity was assessed using an adaptive speech discrimination task, while production variability was determined using a pseudo-word reading task. Analyses of the production data were carried out to quantify average within-phoneme variability, as well as average between-phoneme contrasts. Results show that individuals not only vary in their production and perceptual abilities, but that better discriminators have more distinctive vowel production targets—that is, targets with less within-phoneme variability and greater between-phoneme distances—confirming the initial hypothesis. This association between speech production and perception did not depend on local phoneme density in vowel space. This study suggests that better auditory acuity leads to more precise speech production targets, which may be a consequence of auditory feedback affecting speech production over time. © *2017 Acoustical Society of America.* https://doi.org/10.1121/1.5006899

[CYE] Pages: 2007–2018

I. INTRODUCTION

How do speech perception and speech production interact? Several lines of research have shown that speech production and speech perception are not independent processes, but interact in complicated ways. Investigations of these perception-production interactions can largely be placed in two categories. The first type focuses on short-term effects of perception on production. For example, when a speaker's auditory feedback is manipulated or distorted, his or her speech production is affected (Elman, 1981; Fairbanks and Guttman, 1958; Houde and Jordan, 1998; Purcell and Munhall, 2006). For example, when auditory feedback is delayed by just 200 ms, speakers make more speech errors (Fairbanks and Guttman, 1958), and, when the pitch of individuals' speech is artificially shifted up in auditory feedback, speakers compensate by shifting their pitch downward (Burnett *et al.*, 1998). Although these studies have shown that auditory feedback is not strictly necessary for regular speech production (Lane and Webster, 1991), they also demonstrate that the perception and production systems interact in real time.

The second line of research into the perception-production link focuses on longer-term interactions between speech production and perception, usually by studying correlations between the two. Here, the guiding hypothesis is that if production and perception interact on a daily basis, this will lead to co-variation across individuals. For example, Newman (2003) investigated correlations between acoustic measures of listeners' perceptual prototypes for a given speech category and their average production of members of that category. People whose perceptual prototype of stop consonants had a longer voice onset time (VOT) also tended to produce these consonants with longer VOT.

[a] Portions of this work were presented in "Assessing the link between speech perception and production through individual differences," Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, United Kingdom, August 2015.

[b] Current address: Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Electronic mail: matthias.franken@ugent.be

Another example of research into longer-term interactions concerns studies that have shown a correlation between auditory acuity and vowel production (Perkell *et al.*, 2004; Perkell *et al.*, 2008). In these studies, participants carried out two tasks: (1) a discrimination task on a vowel continuum and (2) an overt reading task. The results showed that participants who were better at the discrimination task produced vowels more consistently (less within-phoneme variability), but spaced them further apart in vowel space (larger between-phoneme acoustic distance). The authors interpret their findings as follows: Better auditory acuity is reflective of more precise speech targets (e.g., smaller target regions in acoustic space), which, in turn, leads to more consistent speech production, as a smaller target region would result in more rejections of non-prototypical productions as "speech errors." A related study is reported by Villacorta *et al.* (2007), who showed that people with higher auditory acuity compensate more strongly in response to altered auditory feedback.

The interplay between speech production and speech perception has also been corroborated in neurobiological studies. Several studies have shown that auditory input is processed differently during speech production compared to passive listening (Christoffels *et al.*, 2011; Franken *et al.*, 2015; Heinks-Maldonado *et al.*, 2006; Houde *et al.*, 2002). Both behaviorally and neurobiologically, it is well established that unexpected auditory feedback leads to subsequent changes in speech production (Behroozmand *et al.*, 2015; Behroozmand*et al.*, 2009; Parkinson *et al.*, 2013). Note, however, that the amount and significance of individual variability in these interactions is not well understood. Along with previous studies, the current study will offer an example of how studying individual variability can illuminate the interplay between perception and production in speech motor control.

Several current theories of speech motor control hypothesize that speech perception contributes to speech production through an auditory feedback mechanism that informs speech motor control (Hickok *et al.*, 2011; Houde and Nagarajan, 2011; Tourville and Guenther, 2011). Further highlighting the important link between perception and production, these models posit that speech production goals are ultimately perceptual targets. In other words, the goal of the speech production process is to produce a particular sound sequence. It is assumed that these sound representations are first acquired via speech perception, making it conceivable that speech production targets will co-vary with individual variability in speech perception. In addition, the speech production process might be tuned over time by perception: Auditory feedback processing may reject the produced speech sound as a deviation from the prototypical representation, leading to compensatory responses. In such a system, individuals with higher perceptual acuity may be more sensitive to speech production that deviates from expected targets. Such deviations might be detected as speech errors, which over time would drive the production system to be more precise (i.e., less variable).

Although the above models make clear predictions about within-phoneme co-variation, it remains unclear whether the production-perception co-variation would also vary across phonemes. It is well established that vowel space is perceptually warped by the presence of phonemes (Kuhl, 1991; Kuhl *et al.*, 2008). Therefore, it is conceivable that associations between speech perception and production may vary both locally, for example, depending on the local phoneme density, as well as cross-linguistically, depending on the language's phoneme inventory. An example of local and cross-linguistic differences in phoneme inventories is shown in Fig. 1, which depicts the vowel inventories of Dutch and English, two closely related languages. Although the two languages have a similar number of vowels, it can be seen in Fig. 1 that in Dutch "front" vowels (those at the higher end of the *F*2 scale) exist in a higher density space than the "back" vowels, whereas this is not the case in English. This is corroborated by analyses of Dutch interphonemic distances in Fig. 1 which showed that, for example, Dutch /ɑ/ lies in a less dense space compared to Dutch /ɛ/, both globally (overall average distance to other phonemes, /ɑ/: 732 ΔHz, /ɛ/: 590 ΔHz) and locally (average distance to three closest phonemes, /ɑ/: 424 ΔHz, /ɛ/: 244 ΔHz; distance to closest phoneme, /ɑ/: 330 ΔHz, /ɛ/: 199 ΔHz). When other phonemes are nearby, it would pay off to have very precise articulatory targets, so that the produced vowel is not confused with the neighboring phonemes. Previous research suggests that neighboring phonemes indeed have an effect on phonemes' target regions, as auditory feedback control is modulated by the presence of nearby phoneme categories (Niziolek and Guenther, 2013). It has not been shown, however, whether phoneme density also affects longer-term interactions between the perception and production systems. For example, higher phoneme density might drive the system to develop stronger perception-production links than those in lower-density regions of the acoustic space. So in denser parts of vowel space, people might be more sensitive to deviations, which would lead them to develop smaller targets. Therefore, we tested the hypothesis that the relationship of speech perception with speech production variability is affected by local phoneme density.

In the present study we address whether the longer-term production-perception interactions discussed above result in associations between perception and production behavior. More specifically, we determine whether auditory acuity, as measured by a speech discrimination task, would be associated with individual variability in vowel productions. This was done by having participants carry out a speech discrimination task and a speech production task, and investigating possible correlations of individual variability across tasks, using a similar paradigm to Perkell *et al.* (2008). In addition, we investigated whether these perception-production associations depend on local vowel density by comparing a pair of front vowels with a pair of back vowels (the bold labels in Fig. 1). In terms of speech discrimination, like Perkell *et al.* (2008) we used a four-interval two-alternative forced choice task, which has been shown to capture lower-level auditory discrimination with relatively little influence from phonemic categories (Gerrits and Schouten, 2004). However, in the present study we measured auditory acuity using a discrimination score, a measure that takes into account both participants' overall discrimination ability, as well as the

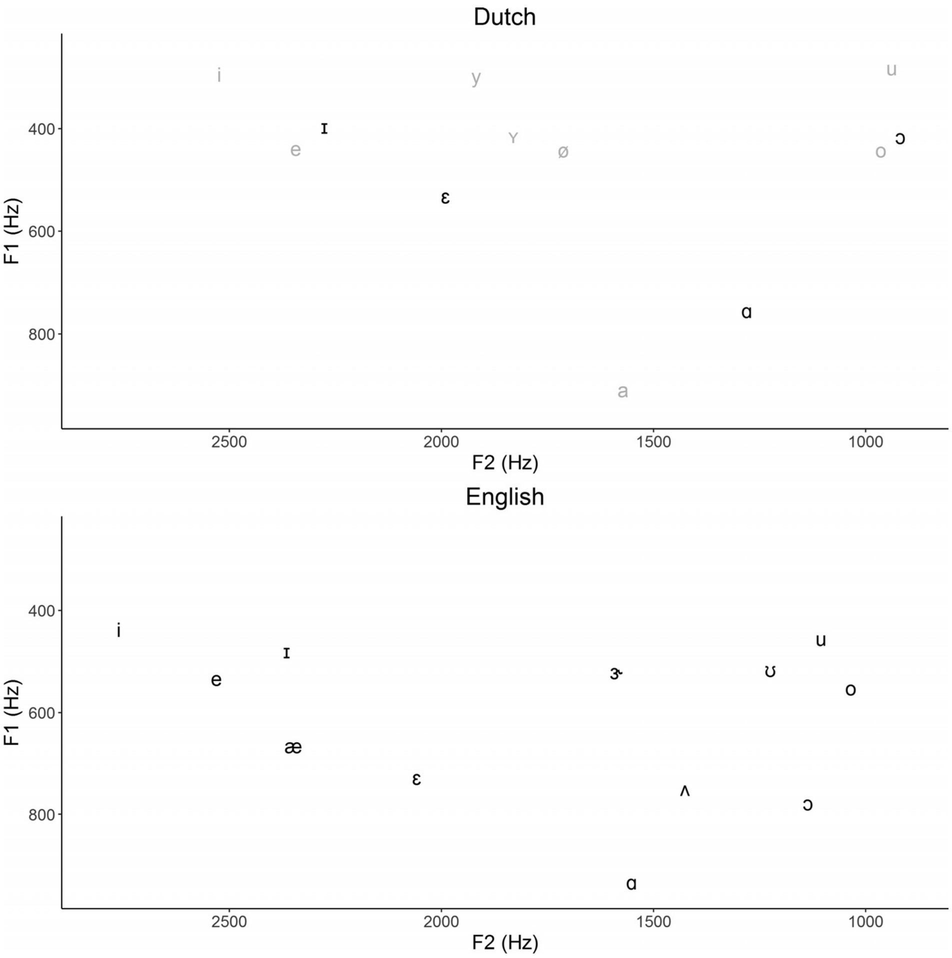

FIG. 1. The vowel spaces of Dutch and English, exemplified by vowels plotted as a function of average first ($F1$) and second ($F2$) formant values. Dutch data shown here are the acoustic values of vowels spoken by females speaking Northern Standard Dutch, reported in Adank *et al.* (2004), excluding the three Dutch diphthongs /ɛɪ/, /ɑu/, and /Œy/. Bold labels indicate the vowels we used in our study (/ɪ/, /ɛ/, /ɑ/, /ɔ/). English data are acoustic measurements of female speakers of American English, taken from Hillenbrand *et al.* (1995).

consistency of their performance, whereas previous methods only captured participants' average performance. The speech production task in the present study was a non-word reading task. In order to characterize production variability, measures were used that take into account distributional properties of vowel space, as well as measures that capture psychophysical properties of speech perception. For example, in addition to characterizing vowel production in terms of $F1$ and $F2$ measurements (as done in most research in this area), we also characterized vowel production in terms of so-called mel-frequency cepstral coefficients (MFCCs). These coefficients represent spectral properties of speech and are widely used in the field of automatic speech recognition. In contrast to $F1$/$F2$ values, they provide a broader representation of the spectral shape of speech sounds and are designed to better reflect the psychophysics of human vowel perception. Compared to the acoustic measures used in Perkell *et al.* (2008), which rely on Euclidian distances in two-dimensional (2-D) vowel space, we believe these measures are better able to capture human speech perception.

If the models of speech production mentioned earlier are correct, long-term interactions between perception and production should lead to co-variability across individuals, and thus we expect our perception measures to correlate with speech production variability. More specifically, these models predict that individuals with better auditory acuity would have more precise vowel targets, and therefore show less production variability. In addition, we investigate whether these predicted associations between perception and production vary as a function of local phoneme density. Therefore, we will compare higher density Dutch front vowels /ɪ/ and /ɛ/ (from a denser part of vowel space) with Dutch back vowels /ɑ/ and /ɔ/ (from a sparser part of vowel space).

II. METHODS

A. Subjects

Forty healthy volunteers [age: $M = 20$ years old, standard deviation (SD) = 2.2; 24 females] participated after providing written informed consent in accordance with the Declaration of Helsinki and the local ethics committee (the Social Sciences Ethical Committee of Radboud University). All participants had normal hearing, were native speakers of Dutch, and had no history of speech and/or language pathology. Three reported being raised multilingually, and the others were raised monolingually in Dutch (though seven reported speaking a local dialect). All participants also reported how many languages they learned (at school or elsewhere) aside from Dutch. As is common in the Netherlands, most of them reported having learned three languages besides Dutch ($M = 3$, SD = 0.92, range = 2–5).

B. Stimuli

For the discrimination task, two speech continua were created based on recordings of the pseudowords *skef* and *skaf*, spoken by a male native Dutch speaker. From each of these recordings, the two continua (/skɛf/-/skɪf/ and /skɑf/-/skɔf/) were made by manipulating $F1$ and $F2$ values. First, the vowels were excised from each recording. Using Burg's linear predictive coding (LPC) framework, a filter model was obtained by estimating five formants between 0 and 5000 Hz. A source model was obtained using eight prediction coefficients. A number of filter models were created by changing $F1/F2$ values in a stepwise manner, and the endpoints of the continua were based on the average $F1$ and $F2$ values for a male Dutch speaker (Adank*et al.*, 2004), as these came close to the values of the original recording. For the *skaf-skof* continuum, 1001 steps were used (as in Perkell *et al.*, 2008), each one having a change of -0.176 Hz in $F1$ and -0.351 Hz in $F2$. For the *skef-skif* continuum, 543 steps were created, so the Euclidian distance in $F1$-$F2$ space between successive steps was similar to the first continuum ($F1$ change was -0.210 Hz, $F2$ change was 0.332 Hz). This allowed us to compare results on both continua. These filter models were combined with the source model. The results were lowpass-filtered at 2000 Hz and combined with the band-pass filtered original signal (2000 Hz–6000 Hz). This way, it was ensured that above 2000 Hz, the signal was exactly the same as the original. All vowels were manipulated so their average intensity matched that of the original sounds. Finally, the vowels were embedded in the *sk_f* context, which was exactly the same for all stimuli in a continuum (the consonantal frame was taken from the original pseudoword recording).

For the production task, pseudowords were created using a $C_1V_1C_1C_1V_1C_2$ structure, where C_1 could be either one of /k/, /p/, or /t/, V_1 either one of /ɛ/, /ɪ/, /ɑ/, or /ɔ/, and C_2 either one of /p/, /t/, /k/, /f/, /s/, or /x/. This particular structure was used because monosyllabic structures led to too many existing words (rather than pseudowords), and the various consonants used were all voiceless obstruents, making it easier to later determine vowel onsets and offsets in the recordings. Using all possible combinations of these vowels and consonants resulted in 72 unique pseudowords (e.g., *kekkef*, *poppos*).

C. Procedure

The experiment consisted of two tasks, which were administered in counterbalanced order within a single session with a short break in between.

The discrimination task consisted of a four-interval two-alternative forced choice task (Gerrits and Schouten, 2004) with a staircase technique based on the weighted up-down procedure (Kaernbach, 1991; Levitt, 1971). On every trial subjects heard four auditory stimuli: three standard stimuli and one deviant stimulus. The standard stimuli were always one extreme of the continuum (i.e., three times the same stimulus, *skef* for the *skef-skif* continuum, *skaf* for the *skaf-skof* continuum), while the deviant stimulus varied on a trial-by-trial basis. The deviant stimulus occurred in position two or three, and the participant was instructed to push the left button when he or she thought the deviant was the second stimulus, and to push the right button when he or she thought it was the third stimulus. If the participant responded correctly, the difference between the standard and the deviant in the next trial was decreased, otherwise, it was increased. Participants did not receive feedback on their performance.

The discrimination task was divided into four blocks, which alternated between continua. Every block started with a fairly large interval (250 continuum steps or Euclidian distance in $F1$-$F2$ space of around 98.2 ΔHz between standard and deviant stimulus). "Reversal" trials were trials where subjects gave a correct response after a previous incorrect trial, or vice versa. The block ended after a total of 20 reversal trials. The amount of change in the interval size from trial to trial was initially large (a decrease of 25 steps after a correct trial, an increase of 75 after an incorrect trial), and became smaller after the second reversal trial of a block (a decrease of 10 after a correct trial, an increase of 30 after an incorrect trial). Because the increase in interval size after an incorrect trial was always three times the decrease of the interval size after a correct trial, the interval size should theoretically converge to a threshold interval size where people would give a correct answer on 75% of the cases (Kaernbach, 1991).

The production task was a simple pseudoword reading task. Subjects were instructed to read aloud the pseudowords that appeared on the screen, while trying to maintain a constant, normal volume and making sure stress was placed on the second syllable (which was printed in capitals). Subjects were positioned about 30 cm from the microphone and asked to try to keep this distance throughout. The task consisted of four blocks, each of which presented all 72 pseudowords in randomized order. Every pseudoword was thus repeated four times.

D. Hardware

All recordings were made in a soundproof booth and digitized at 44.1 kHz on one channel using a Sennheiser ME64 microphone (Wedemark, Germany), which was set up in the booth and connected through an Alesis Multimix 6 FX audio mixer (Cumberland, RI) to a Windows computer (Redmond, WA) outside the booth. Auditory stimuli were delivered through the same audio mixer, which was connected to Sennheiser HD280-13 headphones. Stimuli presentation and sound recording times were controlled by the same Windows computer running Neurobehavioral Systems Presentation (Albany, CA).

E. Analysis

1. Perception

For the results from the discrimination task, we calculated a threshold value per block by averaging the interval sizes for the last 16 reversal trials. Subsequently, we took the minimal threshold per continuum for each subject. As another measure of discrimination performance, we quantified the consistency between blocks of the same continuum in the following way: We created a linear mixed effects model with block and continuum as fixed effects, subject as a random effect (with random slopes for block and continuum), and the calculated thresholds as dependent variables. The absolute values of the random slopes for block were taken as a measure of between-block inconsistency. Finally, we also calculated a "discrimination score" by multiplying the between-block inconsistency measure by the minimal threshold value. So the discrimination score could be high either because the participant was not very consistent between blocks, or had a high minimal threshold. In other words, a higher discrimination score corresponds to worse performance on the discrimination task.

We also carried out a correlation analysis between the minimal threshold and between-block inconsistency measures in order to characterize the relationship between these two measures.

2. Production

For all recordings, the beginning and ending of the vowel in the second syllable, which always carried stress, was manually determined. Then the duration and formant values were extracted. Formant values were calculated by averaging over a 40 ms time window at the center of the vowel. Five formants were estimated between 0 and either 5 kHz (males) or 5.5 kHz (females) using an iterative Burg algorithm in Praat (Boersma and Weenink, 2013). Even though in the present study we were only interested in $F1$ and $F2$, estimating five formants tends to give a more reliable result (Boersma and Weenink, 2013). For all further analyses, formant values were converted from Hertz to the Bark scale (which is defined so that the critical bands of human hearing all have the width of one Bark; Zwicker, 1961).

In order to capture subjects' production variability, two different measures were taken. The first was vowel dispersion, or the area of the ellipse described by one SD in both $F1$ and $F2$ for that phoneme. This was calculated using the formula of the area of the ellipse

$$\text{vowel dispersion} = \pi xy.$$

Here, x and y correspond to one SD in $F1$ and $F2$, respectively. This corresponds to what others have called "compactness score" (Kartushina and Frauenfelder, 2013, 2014). Vowel dispersion was calculated per vowel, and the results were averaged across vowels within subjects. The second measure was average vowel spacing (AVS), which was the average Mahalanobis distance between the phoneme's centroid and all neighboring phoneme distributions. This was averaged across all possible vowel pairings (i.e., between /ɪ/ centroid and /ɛ/ distribution, /ɛ/ centroid and /ɪ/ distribution, /ɪ/ centroid and /ɔ/ distribution, etc.). A similar measure was also used by Kartushina and Frauenfelder (2013, 2014). Both dispersion and AVS were calculated in $F1$-$F2$ space.

Similar analyses were conducted using MFCCs (Gold *et al.*, 2011). MFCC representations mimic the workings of the filter bank in the inner ear. MFCC calculations were done in Praat by first performing a filter bank analysis with 12 filters (first filter centered at 100 mel, distance between successive filters 100 mel). Subsequently, the filter values were converted to MFCCs using a discrete cosine transform. Finally, dispersion was quantified as the mean Euclidian distance to the centroid in 12-dimensional space (defined by 12 MFCCs), and AVS as the average pairwise distance between vowel centroids in the 12-dimensional MFCC space.

3. Perception vs production

In order to assess the association between perception and production variability, regression analyses were carried out with discrimination score (as defined in Sec. II E 1) as the dependent variable and the production measures, as well as vowel continuum as the predictors. Data points for which Cook's distance was larger than 0.1 for a particular analysis (indicating high residuals and/or high leverage) were removed from that analysis (on average 3.25% of the data points were removed).

III. RESULTS

A. Discrimination

For every participant, the discrimination threshold was calculated for every block in both continua. The results for a representative participant are shown in Fig. 2. Although the average threshold across subjects for both continua was lower in the second block [/ɛ/-/ɪ/ block 1: $M = 83.1$ Δbark (SD = 53.1), block 2: $M = 72.2$ Δbark (42.4); /ɑ/-/ɔ/ block 1: $M = 128.5$ Δbark (62.9), block 2: $M = 105.8$ Δbark (50.3)], there were also subjects who showed an increased threshold for both continua in the second block (17 subjects for /ɛ/-/ɪ/, 13 subjects for /ɑ/-/ɔ/). If we take the minimum threshold for each participant and each continuum, we see that the /ɑ/-/ɔ/ continuum was harder than the /ɛ/-/ɪ/ continuum [/ɛ/-/ɪ/: $M = 60.2$ Δbark (31.3); /ɑ/-/ɔ/: $M = 97.4$ Δbark (48.3)]. This difference was significant [t (76.71) = −4.86, $p < 0.001$, t-test done on log-transformed threshold values].

With respect to within-subject variability, there were positive correlations between participants' discrimination threshold in the first block and their threshold in the second block for both continua (for /ɛ/-/ɪ/: $r(38) = 0.63$, $p < 0.001$ and for /ɑ/-/ɔ/: $r(38) = 0.59$, $p < 0.001$). Although a positive correlation was expected (given that participants performed the same task on the same stimuli), it explained only about 40% and 36% of the variability, respectively, indicating that participants did not perform consistently in either block. To quantify this variability, we performed a linear mixed effects model analysis on the participants' thresholds with continuum (/ɛ/-/ɪ/ vs /ɑ/-/ɔ/) and block (block 1 or block 2) as fixed effects and random slopes within subjects. The results are shown in Table I.

As a measure of participants' inconsistency in their performance, we took the absolute value of the random effects for the block predictor. For both continua, this inconsistency value correlated weakly with the participants' minimal thresholds [/ɛ/-/ɪ/: $r(37) = 0.24$; /ɑ/-/ɔ/: $r(37) = 0.23$; see Fig. 3]. In other words, participants with a higher minimal threshold (worse discrimination performance) also performed less consistently in the discrimination task.

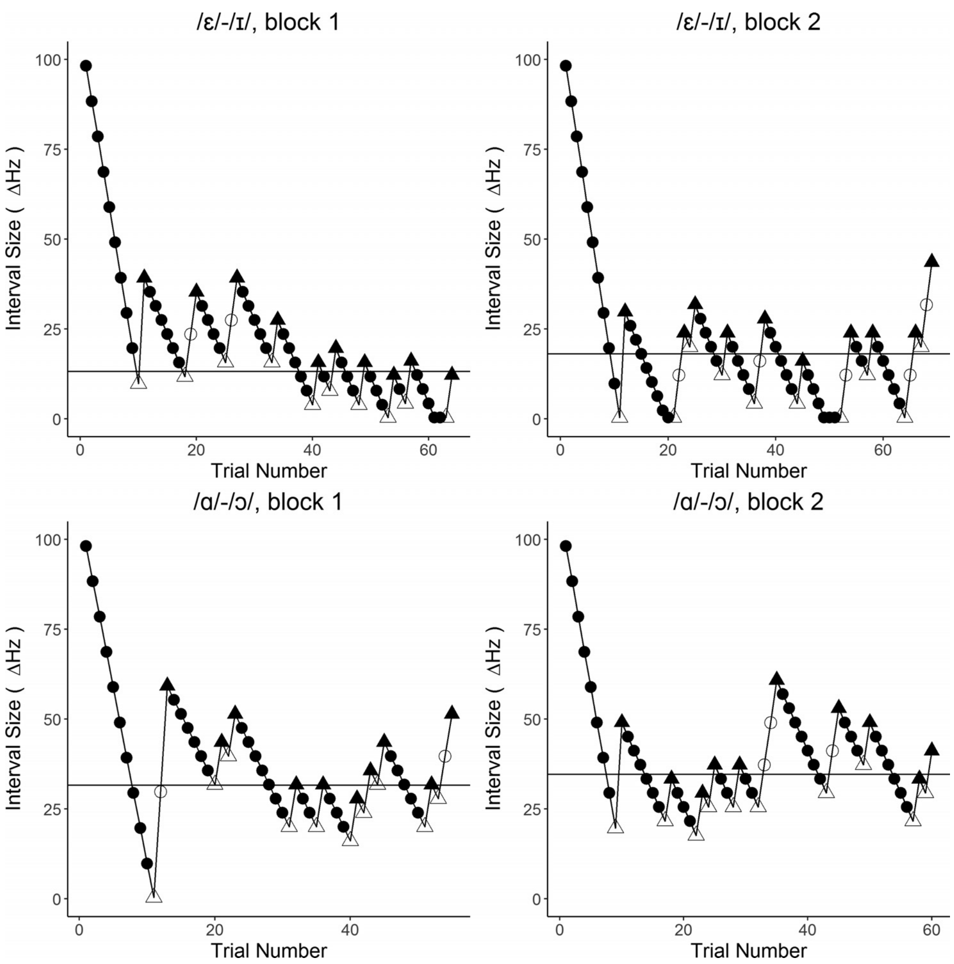

FIG. 2. Discrimination results for a representative participant. Every panel of the plots shows the interval size as a function of trial number for a particular experimental block. The top row shows the two blocks for the /ɛ/-/ɪ/ continuum, the bottom row shows those for the /ɑ/-/ɔ/ continuum. The left column shows the first block for each continuum and the right column shows the second block. The solid symbols indicate trials that were answered correctly; empty symbols indicate trials in which the response was incorrect. Triangles indicate reversal trials. The horizontal line indicates the threshold calculated for that block.

In order to quantify both performance inconsistency and discrimination threshold, we used the discrimination score (inconsistency*threshold) in subsequent analyses.

B. Production

To quantify speech production variability, we used measures in $F1$-$F2$ space, as the majority of research in acoustic phonetics characterizes vowel acoustically in terms of formant values. We also used measures in MFCC space. MFCC values are designed to capture vowel acoustics in a way that is closer to human perception, as these coefficients are based on filter banks similar to known variation of the ear's critical bandwidths (Davis and Mermelstein, 1980). In both $F1$-$F2$ and MFCC domains we had a measure of within-phoneme variability and a measure of between-phoneme distance.

TABLE I. Linear mixed effects model results, looking at discrimination thresholds in terms of block and continuum, with random slopes for both within subjects.

	Estimates	t-values		Estimates
Fixed effects			Random effects	
Intercept	86.002 (7.998)	10.752	Intercept (subjects)	2037.4
Block	−16.736 (6.000)	−2.789	Block (subjects)	744.6
Continuum	39.508 (8.458)	4.671	Continuum (subjects)	2166.3
Residual	695.4			

In $F1$–$F2$ space, the within-phoneme variability measure (ellipse area; see Fig. 4 for an example) had cross-participant means of 0.27 Δbark (0.17) for /ɑ/, 0.25 Δbark (0.10) for /ɛ/, 0.17 Δbark (0.09) for /ɪ/ and 0.30 Δbark (0.21) for /ɔ/ (SDs between brackets). For the between-phoneme distance measure (AVS or mean squared Mahalanobis distances), we find a mean of 160.9 Δbark^2 (70.2). For further correlation analyses (see Sec. III C), variability due to gender was removed from this measure, as this also affects AVS values. This was done by generating a linear model with AVS as the independent measure and a single predictor that coded for gender. The linear model showed that male speakers had smaller AVS values (i.e., a smaller vowel space) than female speakers [$F(1,38) = 12.98$, $p < 0.001$], as is well known from the literature (Simpson, 2001, 2009). The residuals of this linear model, reflecting variability in AVS that cannot be attributed to gender differences, were used as input in the correlation analyses.

C. Production-perception associations

Regression analyses were performed in order to compare individual variability in the discrimination and production tasks. Specifically, participants' perceptual discrimination scores were used as the dependent variable with vowel continuum (/ɛ/-/ɪ/ and /ɑ/-/ɔ/) as a predictor and two production-based predictors: dispersion (within-phoneme variability; vowel ellipse area) and AVS (between-phoneme distance).

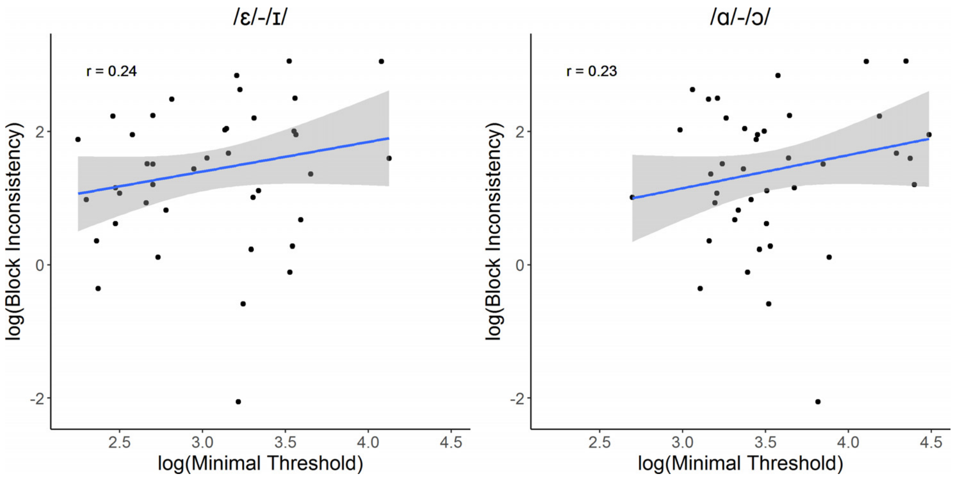

FIG. 3. (Color online) Scatter plots of the association between participants' minimal threshold and their block inconsistency score (both log-transformed) for the /ɛ/-/ɪ/ continuum (left) and for the /ɑ/-/ɔ/ continuum (right). The superimposed line represents the best linear fit, ignoring the outlier at the low end of the block inconsistency measure. Shading represents the 95% confidence interval.

These analyses were run twice, once with the $F1$–$F2$ production measures and once with the MFCC measures.

For the production measures in $F1$–$F2$ space, we first ran a full model, including the predictors dispersion, AVS, and vowel continuum, as well as the latter's interaction terms with both production measures. The results of the regression analysis (after having removed four data points for which Cook's distance was over 0.1) are shown in Table II.

As is shown in Table II, none of the interaction terms is significant, showing that any association between the production and the perception measures is not dependent on the vowel. Next, the same variables were entered in a stepwise regression procedure. This procedure allowed us to arrive at a model in which predictors that do not significantly increase the model's goodness of fit were left out. The outcome of this procedure suggested the best final model included both production terms (dispersion and AVS) as predictors, excluding vowel continuum, as well as the interaction terms. The results of this final model are shown in Table III.

The results show a significant (negative) main effect of AVS, suggesting that better performance in the discrimination task (i.e., lower discrimination score) was associated with larger between-phoneme distances in production. In other words, people who were better in speech discrimination produced vowels that were spaced further apart in vowel space. In addition, the main effect of vowel dispersion is marginally significant, indicating that people who produce vowels with less within-phoneme variability perform better at the discrimination task. However, this should be interpreted with caution given that this effect is only marginally significant in the final model, and not significant in the full model (see Table II). The absence of interaction terms between any production measure and vowel continuum confirms that the association between production and performance in the discrimination task does not depend on specific vowels.

The pattern of results as shown in the regression analyses is in line with the pairwise correlation analyses per vowel continuum, shown in Fig. 5. We found positive correlation coefficients for the comparison between vowel dispersion (i.e., within-phoneme variability) and discrimination score [/ɑ/-/ɔ/: $r(36) = 0.38$, $p = 0.02$, Fig. 5 top right and /ɛ/-/ɪ/: $r(36) = 0.25$, $p = 0.14$, Fig. 5 top left]. This confirms the finding of the regression analyses that better discrimination performance (i.e., a lower discrimination score) was associated with less within-phoneme variability, that is, more precise vowel production. For the between-phoneme distance measure, or AVS, we found negative correlation coefficients

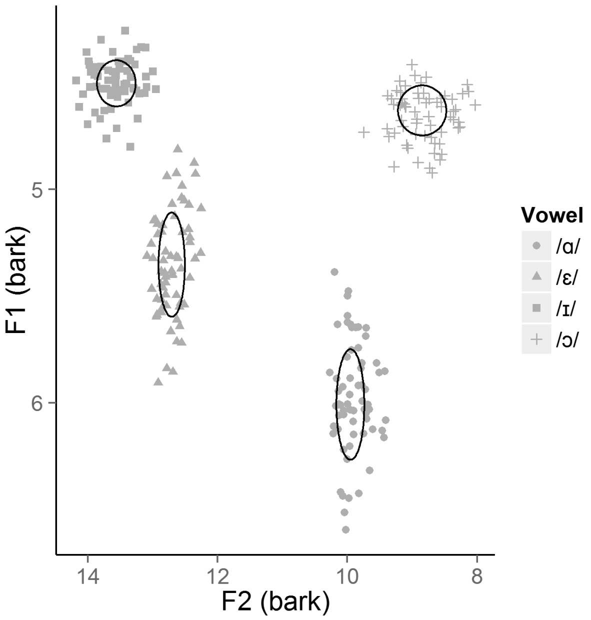

FIG. 4. Production data from a representative participant. The grey symbols show single trial results in terms of $F1$ and $F2$ (both in bark). Symbol shapes indicate the vowel and ellipses show the within-phoneme variability measure (area of the ellipse) for each vowel phoneme.

TABLE II. Regression coefficients for the full model with production measures in $F1$-$F2$ space. Asterisks indicate significance at the 0.05 alpha level.

	Estimates	*t*-values	*p*-values
Intercept	9.94 (1.47)	6.77	<0.0001*
AVS	−0.71 (0.30)	−2.39	0.019*
Dispersion	0.31 (0.67)	0.46	0.65
Vowel continuum	1.74 (2.25)	0.77	0.44
AVS: vowel continuum	−0.11 (0.41)	−0.27	0.78
Dispersion: vowel continuum	0.62 (0.91)	0.68	0.50

TABLE III. Regression coefficients for the final model with production measures in $F1$-$F2$ space. Asterisks indicate significance at the 0.05 alpha level.

	Estimates	t-values	p-values
Intercept	10.97 (1.06)	10.33	<0.0001*
AVS	−0.73 (0.20)	−3.73	0.00038*
Dispersion	0.79 (0.40)	1.96	0.053

[for /ɑ/-/ɔ/: $r(37) = -0.33$, $p = 0.04$, Fig. 5 bottom right, and for /ɛ/-/ɪ/: $r(37) = -0.19$, $p = 0.25$, Fig. 5 bottom left]. Again, this confirms the regression results, showing that better discrimination performance was associated with larger between-phoneme distances. In other words, speakers who were better at the discrimination task produced vowels that were further apart in vowel space. Although not all significant, the scatter plots in Fig. 5 are similar across vowel continua (and the correlations have the same direction), which is in line with the lack of interactions with vowel continuum in the regression analyses being significant. As with the regression analyses, these results are consistent with an association between speech perception and production that does not dependent on the vowel (and thus on local vowel density in vowel space).

Note that in the analyses reported here, the production measures were calculated across the entire vowel space, in contrast to the discrimination performance (which was continuum specific). This was done to get more reliable estimates of people's phoneme dispersion and average between-phoneme spacing. Production variability can be affected by various factors, and averaging across phonemes generates, in our view, a better estimate of overall within-phoneme variability and between-phoneme distinctions. Consistent with the assumption that the production data for individual phonemes are more variable, there is no significant association for vowel dispersion [(/ɑ/-/ɔ/: $r(36) = 0.20$, $p = 0.23$, and /ɛ/-/ɪ/: $r(36) = 0.20$, $p = 0.24$], or for AVS [/ɑ/-/ɔ/: $r(36) = -0.15$, $p = 0.36$, and /ɛ/-/ɪ/: $r(36) = 0.20$, $p = 0.22$] if vowel dispersion and AVS are computed separately for each continuum. Note that, for dispersion, this was done by averaging the dispersion values for the two endpoints of each continuum to correspond with the discrimination measures because they necessarily involve both vowels.

Similar results were found when using the production measures in MFCC space. Table IV shows the results of a full regression model, including MFCC production measures dispersion and AVS, as well as the vowel continuum term and its interactions with the production measures (after having removed two data points for which Cook's distance was over 0.1).

Similar to the results above, none of the interaction terms with vowel continuum is significant. Next, the same variables were entered in a stepwise regression procedure. The outcome of the stepwise regression suggested that the best final model included both production terms (dispersion and AVS) as predictors, as well as vowel continuum, but excluding the interaction terms. The results of this final model are shown in Table V.

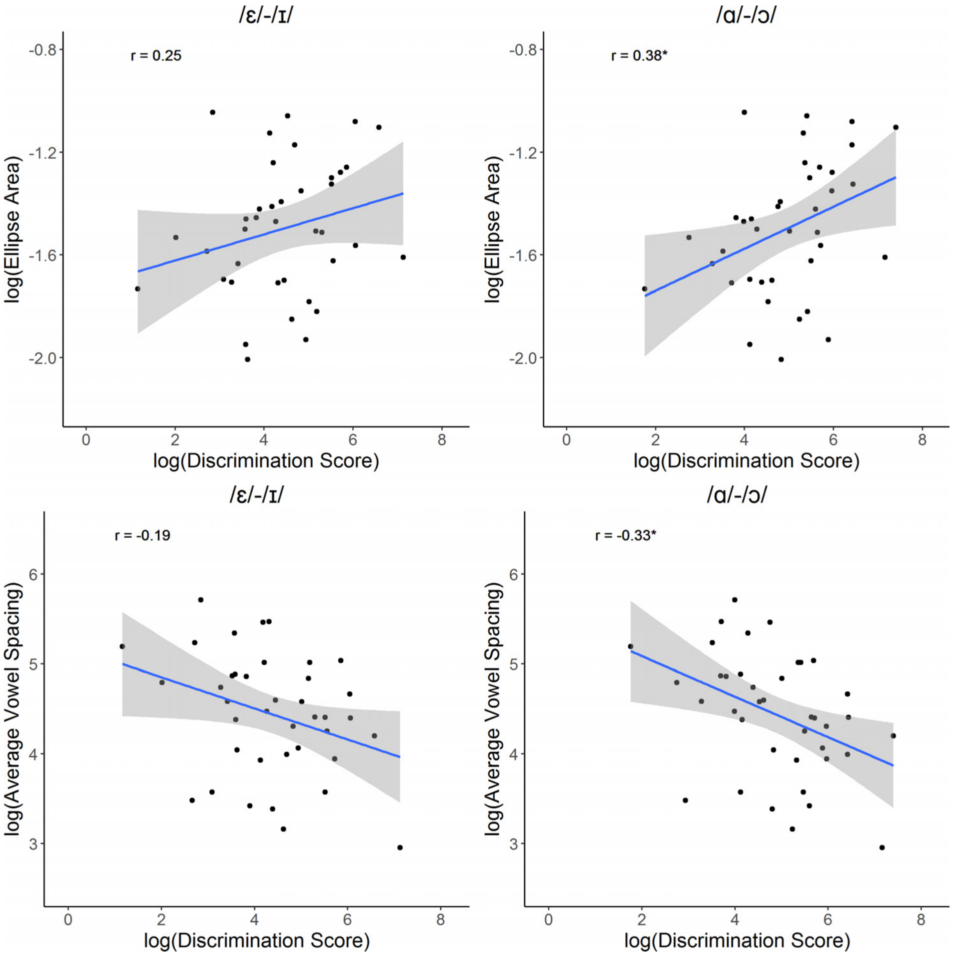

FIG. 5. (Color online) Scatter plots of correlation analyses in $F1$-$F2$ space. Top row shows comparisons between discrimination score (x axis) and within-phoneme variability (y axis). Bottom row shows comparisons between discrimination score and average between-phoneme distance (y axis). Left column shows results for the /ɛ/-/ɪ/ continuum and the right column shows results for the /ɑ/-/ɔ/ continuum. Superimposed lines represent the best linear fit. Shading represents the 95% confidence interval.

TABLE IV. Regression coefficients for the full model with production measures in MFCC space. Asterisks indicate significance at the 0.05 alpha level.

	Estimates	*t*-values	*p*-values
Intercept	10.59 (14.66)	0.72	0.47
AVS	−5.36 (2.47)	−2.17	0.033*
Dispersion	5.61 (1.98)	2.83	0.006*
Vowel continuum	12.74 (20.73)	0.62	0.54
AVS: Vowel continuum	−1.12 (3.50)	−0.32	0.75
Dispersion: Vowel continuum	−1.37 (2.80)	−0.49	0.63

The results of the final model show significant main effects of AVS, dispersion and vowel continuum. The main effects of AVS and dispersion confirm the effects already found in the full model (Table IV). These effects suggest that better auditory speech discrimination is associated with smaller within-phoneme variability (less dispersion) in production, as well as larger between-phoneme distances. In other words, better discriminators produce vowels more precisely and space them further apart in vowel space. In addition, we find a significant main effect of vowel continuum, suggesting discrimination performance is worse in the /ɑ/-/ɔ/ continuum compared to the /ɛ/-/ɪ/ continuum. Although this effect should be taken with some caution, as this was not significant in the full model, it is in line with the significant difference between /ɛ/-/ɪ/ and /ɑ/-/ɔ/ discrimination performance found earlier.

TABLE V. Regression coefficients for the final model with production measures in MFCC space. Asterisks indicate significance at the 0.05 alpha level.

	Estimates	*t*-values	*p*-values
Intercept	16.70 (10.25)	1.63	0.11
AVS	−5.92 (1.73)	−3.42	0.0010*
Dispersion	4.92 (1.39)	3.55	0.00067*
Vowel_Continuum	0.53 (0.25)	2.11	0.039*

The same pattern of results can be seen in the pairwise correlation analyses per vowel continua, shown in Fig. 6. For the within-phoneme variability measure, we found a significant positive correlation for the /ɛ/-/ɪ/ continuum [$r(37) = 0.35$, $p = 0.03$, see Fig. 6 top left], but not for the /ɑ/-/ɔ/ continuum [$r(37) = 0.25$, $p = 0.12$, see Fig. 6 top right]. The between-phoneme distance measures in MFCC space also showed negative correlation coefficients, which was significant for the /ɑ/-/ɔ/ continuum [$r(37) = -0.34$, $p = 0.03$, see Fig. 6 bottom right], but not for the /ɛ/-/ɪ/ continuum [$r(37) = -0.24$, $p = 0.15$, see Fig. 6 bottom left]. These pairwise correlations are consistent with the results of the regression analyses and the results from the analyses of the $F1$-$F2$ space measures. The absence of any significant interaction between vowel continuum and the production measures shows that these kinds of production-perception association are not dependent on specific vowels. In line with this, the scatter plots in Fig. 6 look similar across vowel

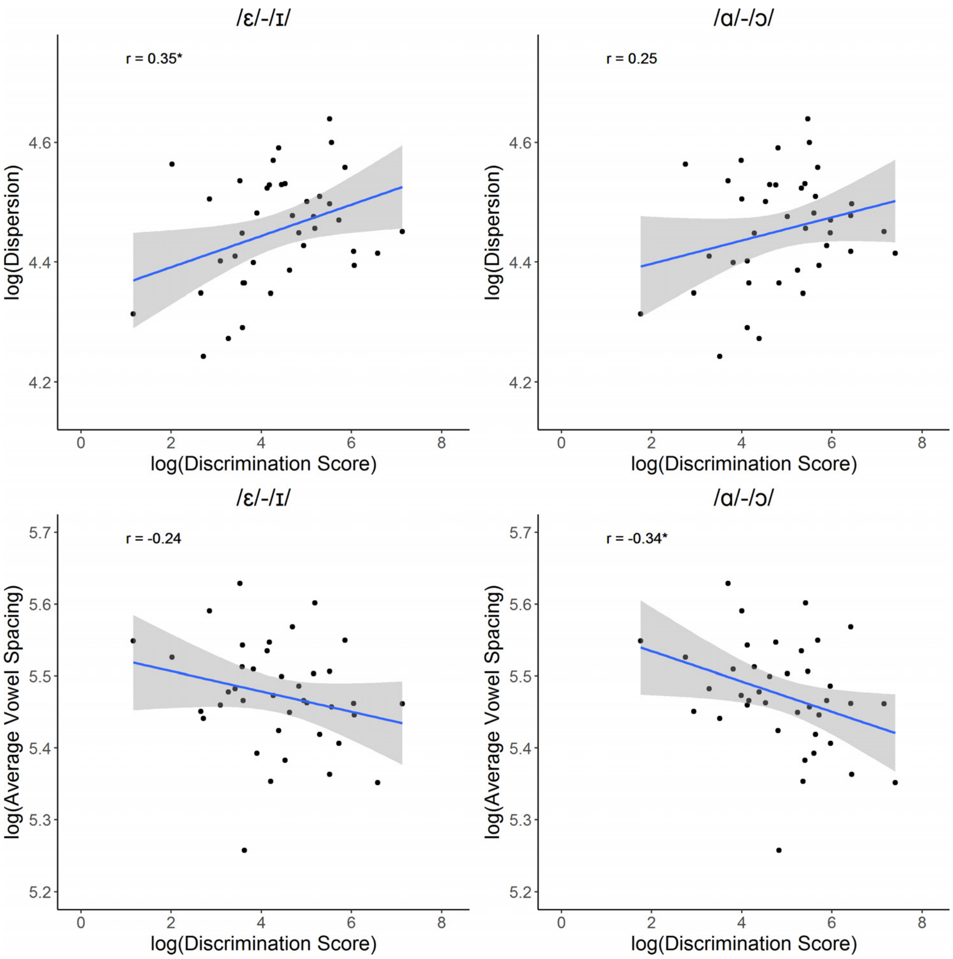

FIG. 6. (Color online) Scatter plots of correlation analyses in MFCC space. The top row shows comparisons between discrimination score (x axis) and within-phoneme variability (y axis). The bottom row shows comparisons between discrimination score and average between-phoneme distance (y axis). Left column shows results for the /ɛ/-/ɪ/ continuum, the right column for the /ɑ/-/ɔ/ continuum. Superimposed lines represent the best linear fit, shading represents 95% confidence interval.

continua, with correlations in the same direction and of similar magnitudes.

In order to be able to compare the present results with those reported in Perkell *et al.* (2008), analyses were also performed with the methods used in that study (see the Appendix). Associations were reported in that study between auditory acuity and both vowel dispersion and AVS. Using these analysis methods on the present data did not show statistically significant correlations, suggesting that the methods presented here (using among others metrics based on ellipse area, Mahalanobis distances, and MFCCs) were better able to capture these perception-production associations. The direction of the (non-significant) correlation trends was, however, in the same direction. That is, people with higher auditory acuity tended to show less vowel dispersion and more AVS.

IV. DISCUSSION

In the present study we compared individual variability in a speech perception task with variability in speech production. Our reasoning was that if, as the literature suggests, speech production and speech perception interact over time, individual differences in these domains should correlate. The results showed that better discrimination performance was associated with less within-phoneme variability in production, as well as with larger average between-phoneme distances. This picture emerged both from the analyses using production measures in $F1$-$F2$ space, as well as from the analyses using measures in MFCC space. In addition, none of the regression analyses showed a significant interaction with vowel continuum regardless of whether the target vowels were in the denser front part of vowel space or in the sparser back part. This suggests that the perception-production association was not dependent on specific vowels or local phoneme density, but instead holds across vowel space. This is also corroborated by the fact that when we used continuum-specific production measures, no significant associations were found.

These results are largely in line with previous findings by Perkell *et al.* (2004) and Perkell *et al.* (2008), although these earlier studies reported much stronger effects. It is unclear what drives the difference in effect sizes. Although we tested native speakers of Dutch whereas Perkell *et al.* (2004) and Perkell *et al.* (2008) tested native speakers of English, we would not expect the link between perception and production, in general, to be dependent on native language.

In addition to differences in language, and hence phoneme space, there were other differences between our study and Perkell *et al.* (2008), as we used different measures to quantify perception and production variability. In our discrimination task, we noticed a fairly high amount of variability between blocks within the same subject and the same continuum. This drove us to use the measure we called the discrimination score, which captured both the participants' best discrimination performance, as well as their consistency across blocks. Perkell *et al.* (2008) did not report on the variability of discrimination performance within subjects, and simply used the measure of the participants' discrimination threshold. In terms of production measures, we have used measures that should take into account vowel distributions and perceptual warping of acoustic space to a larger degree. With respect to the measures in $F1$-$F2$ space, we used the area of the ellipse and Mahalanobis distance, whereas Perkell *et al.* (2008) used the SD of the distribution and Euclidian distance. The measures used in our study, proposed earlier by Kartushina and Frauenfelder (2013, 2014), take into account differential distribution shapes of the different phonemes, and therefore are likely to better reflect phoneme variability. Additionally, we characterized vowels in terms of MFCCs, which imitate the transfer function of the cochlea in the human ear, thus capturing the vowels' acoustics in a way similar to the human ear. Although all these differences between the current study and the study by Perkell *et al.* (2008) may have contributed to the differences in effect size, note that using the same methods as Perkell *et al.* (2008) on the current data did not yield larger effect sizes.

The within-phonemic variability as measured by our vowel dispersion metric may capture both speech target precision, as well as variability due to coarticulation across consonantal contexts. In order to estimate the effect of coarticulatory variability, we recalculated the dispersion measure after having removed the variance due to phonological context. Correlation analyses with discrimination scores led to similar results [$r(36) = 0.23$ for /ɛ/-/ɪ/, where it was 0.24 and $r(36) = 0.26$ for /ɑ/-/ɔ/, where it was 0.37]. Thus, for the /ɑ/-/ɔ/ continuum, it seems at least part of the association may be driven by coarticulatory variability. This would suggest that at least for this continuum, better discriminators show less coarticulatory variability. If we adopt a view on coarticulation where the phonological context affects an underlying phonemic target (Farnetani and Recasens, 2010), this is still in line with the hypothesis that the underlying target region is more precise (or more robust) for better discriminators. A more robust underlying target region would then leave less room for coarticulation effects to take place, and thus less coarticulatory variability.

The overall consistency of our results with those of Perkell *et al.* (2004) and Perkell *et al.* (2008) nevertheless shows that people with better auditory acuity have reduced production variability. These findings are, in turn, consistent with several recent models of speech production. Many of these models consider the goal of speech production to be at least partially an acoustic goal. Therefore, individual differences in auditory perception may well affect variability in speech production targets. One example of these models is presented in Perkell (2007, 2012), where speech production targets are explicitly considered to be regions in auditory space. According to this view, better auditory discrimination performance corresponds to having a higher resolution in auditory space, which, in turn, leads to more precise auditory speech production targets and therefore to more precise or less variable speech production.

Another important component of recent theoretical frameworks is the interaction between feedforward and feedback control of speech production. In the feedback control part of the system, the auditory target is activated during speech production, thus, enabling comparison with incoming auditory feedback. When mismatches occur between predicted and actual auditory information, corrections can be implemented in real time and, over time, the feedforward mechanisms guiding motor targets can be updated and maintained. There are multiple possible means by which auditory acuity might influence this learning. First, under this sort of model (e.g., Tourville and Guenther, 2011), individuals who are better at discriminating speech sounds would then become better at detecting mismatches between feedback and speech targets, and would therefore update their feedforward mechanisms more readily. The end result of this process playing out over time is that speech productions at the periphery of the target region (in auditory space) would be recognized as an error for some individuals, but not for others. If it is recognized as an error, this may lead, over time, to changes in the feedforward commands as such "errors" should be avoided, effectively decreasing the variability in speech production. Consistent with this mechanism, Villacorta *et al.* (2007) demonstrated that speakers with higher auditory acuity show a greater behavioral response to altered auditory feedback. Thus, a second possibility is that people with better auditory acuity respond more strongly to altered auditory feedback, thus, decreasing the variability in their speech production over time. Finally, with respect to inter-phonemic distances, some studies have suggested previously that mismatches that bring the speech sound closer to a neighboring phoneme are more readily perceived or are compensated for more strongly than mismatches that bring the result farther away from a neighboring phoneme (Lametti *et al.*, 2014; Niziolek and Guenther, 2013). Over time, this may lead individuals who are more sensitive to these mismatches to produce speech sounds that are spaced further apart in auditory space, which in this study was quantified as larger AVS. Such a result was attained through simulations with the DIVA model (Tourville and Guenther, 2011; Perkell, 2012).

V. CONCLUSION

In this study, we investigated the association between speech production and speech perception and whether this association would be dependent on local phoneme density. The results show that, overall, speakers with higher auditory acuity produced vowels more distinctively, that is, vowels that were spaced further apart and with less within-category variability. This association did not depend on local phonemic density in vowel space. These findings corroborate current thinking about feedback processing during speech production and the role of auditory information. Furthermore, this study offers insights into individual variability in speech production, which to date is still not well understood. More specifically, our findings are consistent with predictions from current theoretical models of speech motor control, and suggest that speakers with higher auditory acuity have more precise speech production targets, which subsequently shapes their speech production.

ACKNOWLEDGMENTS

This work was partly financed by the Netherlands Organization for Scientific Research (NWO).

APPENDIX

In addition to the results reported in the main text, and in order to be able to compare our results to the previous literature, the data were also analyzed with the methods reported by Perkell *et al.* (2008).

For the perception metric, the discrimination threshold was estimated for every block, as described in the main text. Subsequently, auditory acuity was estimated as the inverse of the discrimination threshold and averaged across blocks for each participant. With respect to the production data, all formant values were converted to mels using the following formula (Boersma and Weenink, 2013):

$$m = 1127 \log\left(1 + \frac{f}{700}\right).$$

Here, f is the formant estimate in Hertz and m is the estimate in mels. Dispersion was calculated for each phoneme as the average Euclidian distance to the centroid in $F1$-$F2$ space, and subsequently averaged across phonemes. AVS was defined as the using Euclidian distance in mel between the centroids of all possible vowel pairs, and subsequently these values were averaged across vowel pairs.

Correlation tests reported no significant correlation between acuity and dispersion ($r = -0.14$), nor between acuity and AVS ($r = 0.20$). Note though, that the direction of the trend corresponds to the results reported in the main text (the sign of the correlation coefficients is different as the acuity is the inverse of the discrimination threshold).

Adank, P., van Hout, R., and Smits, R. (**2004**). "An acoustic description of the vowels of Northern and Southern Standard Dutch," J. Acoust. Soc. Am. **116**(3), 1729–1738.

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., and Larson, C. R. (**2015**). "Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control," Front. Neurosci. **9**, 109.

Behroozmand, R., Karvelis, L., Liu, H., and Larson, C. R. (**2009**). "Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation," Clin. Neurophysiol. **120**(7), 1303–1312.

Boersma, P., and Weenink, D. (**2013**). "Praat: Doing phonetics by computer [computer program]," http://www.praat.org (Last viewed November 2, 2014).

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice $F0$ responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**(6), 3153–3161.

Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., and Schiller, N. O. (**2011**). "The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex," PLoS One **6**(5), e18307.

Davis, S., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust. Speech Signal Process. **28**(4), 357–366.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**(1), 45–50.

Fairbanks, G., and Guttman, N. (**1958**). "Effects of delayed auditory-feedback upon articulation," J. Speech Hear. Res. **1**(1), 12–22.

Farnetani, E., and Recasens, D. (**2010**). "Coarticulation and connected speech processes," in *The Handbook of Phonetic Sciences*, 2nd ed., edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon (Blackwell, Oxford, UK), pp. 316–352.

Franken, M. K., Hagoort, P., and Acheson, D. J. (**2015**). "Modulations of the auditory M100 in an imitation task," Brain Lang. **142**, 18–23.

Gerrits, E., and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**(3), 363–376.

Gold, B., Morgan, N., and Ellis, D. (**2011**). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, 2nd ed. (Wiley, Hoboken, NJ).

Heinks-Maldonado, T. H., Nagarajan, S. S., and Houde, J. F. (**2006**). "Magnetoencephalographic evidence for a precise forward model in speech production," Neuroreport **17**(13), 1375–1379.

Hickok, G., Houde, J., and Rong, F. (**2011**). "Sensorimotor integration in speech processing: Computational basis and neural organization," Neuron **69**(3), 407–422.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

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

Houde, J. F., and Nagarajan, S. S. (**2011**). "Speech production as state feedback control," Front. Hum. Neurosci. **5**, 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (**2002**). "Modulation of the auditory cortex during speech: An MEG study," J. Cognit. Neurosci. **14**(8), 1125–1138.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**(3), 227–229.

Kartushina, N., and Frauenfelder, U. H. (**2013**). "Foreign accents and native sloppiness: The role of individual native production on non-native vowel pronunciation," in *Proceedings of Phonetics, Phonology and Language Contact*, Paris, France, pp. 25–28.

Kartushina, N., and Frauenfelder, U. H. (**2014**). "On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation," Front. Psychol. **5**, 1246.

Kuhl, P. K. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," Percept. Psychophys. **50**(2), 93–107.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (**2008**). "Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e)," Philos. Trans. R. Soc. Lond. B Biol. Sci. **363**(1493), 979–1000.

Lametti, D. R., Krol, S. A., Shiller, D. M., and Ostry, D. J. (**2014**). "Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning," Psychol. Sci. **25**(7), 1325–1336.

Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**(2), 859–866.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**(5), 2850–2860.

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.

Parkinson, A. L., Korzyukov, O., Larson, C. R., Litvak, V., and Robin, D. A. (**2013**). "Modulation of effective connectivity during vocalization with perturbed auditory feedback," Neuropsychologia **51**(8), 1471–1480.

Perkell, J. S. (**2007**). "Sensory goals and control mechanisms for phonemic articulations," in *Proceedings of the 16th International Congress of the Phonetic Sciences*, Saarbruecken, Germany, pp. 169–174.

Perkell, J. S. (**2012**). "Movement goals and feedback and feedforward control mechanisms in speech production," J. Neurolinguist. **25**(5), 382–407.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contracts in related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**(4), 2338–2344.

Perkell, J. S., Lane, H., Ghosh, S. S., Matthies, M. L., Tiede, M., Guenther, F. H., and Ménard, L. (**2008**). "Mechanisms of vowel production: Auditory goals and speaker acuity," in *8th International Seminar on Speech Production*, Strasbourg, France, pp. 29–32.

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.

Simpson, A. P. (**2001**). "Dynamic consequences of differences in male and female vocal tract dimensions," J. Acoust. Soc. Am. **109**(5), 2153–2164.

Simpson, A. P. (**2009**). "Phonetic differences between male and female speech," Lang. Linguist. Compass **3**(2), 621–640.

Tourville, J. A., and Guenther, F. H. (**2011**). "The DIVA model: A neural theory of speech acquisition and production," Lang. Cogn. Process. **26**(7), 952–981.

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.

Zwicker, E. (**1961**). "Subdivision of audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**(2), 248.