The Control of Token-to-Token Variability: an Experimental and Modeling Study

Christine Mooshammer[1], Pascal Perrier[2], Susanne Fuchs[1,3], Christian Geng[1] and Yohan Payan[4]
[1]*ZAS – Research Centre for General Linguistics, Jägerstr. 10/11, 10117 Berlin, Germany*
timo@zas.gwz-berlin.de
[2]*Institut de la Communication Parlée, UMR 5009, INPG & Université Stendhal, 46 Avenue Félix Viallet, 38031 Grenoble Cédex 1, France.*
[3]*Queen Margaret University College, Speech & Language Sciences, Clerwood Terrace, Edinburgh EH 12 8TS,UK*
[4]*TIMC Laboratory, Faculté de Médecine, Institut Albert Bonniot, 38706 La Tronche Cedex, France.*

Abstract. The articulatory token-to-token variability in the production of German vowels is investigated with simultaneous EMMA and EPG recordings. The potential role of physical constraints, such as the contacts between tongue and palate measured by EPG, and the biomechanical properties of the tongue, simulated with a 2D finite element model is evaluated. Our results suggest that the control of high front vowels makes use of the palatal contacts, while the variability of low vowels is essentially oriented along the main axis of deformation of the tongue, the high/front-to-low/back direction.

1. Introduction

Speech production involves a succession of non-linear transformations between different physical domains, from the motor control level to the acoustical space, via the articulatory and the vocal tract geometrical spaces. This complexity is at the origin of major debates about the characterization of the speaker task, and about its relation with the perceptual system (Stevens & Blumstein, 1981; Lindblom, 1988; Browman & Goldstein, 1990). Most of the contributions to this issue were focused on the search of an invariant in one of the above mentioned physical domains (for a summary see Perkell & Klatt, 1986). Methodologically, the paradigm often consists in looking for a physical characteristic that would be shared by several productions of a given phoneme in different phonetic contexts, under varying speaking conditions, where speakers are asked to modify speaking rate, stress location or clarity of their elocution. An alternative paradigm was proposed by Perkell and Nelson (1985). It consists in observing the variability of the physical signals associated to the repetitions of the same speaking task, and to study in which domain the variability is the most limited, and which could be the control strategy to do so. This approach is in agreement with a motor control theory recently proposed by Harris and colleagues (Harris, 1998; Harris & Wolpert, 1998). This theory suggests that motor control strategies underlying the production of target directed movements would be determined by the minimization of the end-position variability that is due to the corruption of the control signals by neural noise, from one repetition of the task to the next. Such a theoretical framework incites us to think that, in the line of Perkell & Nelson (1985), studying the variability in different physical domains of speech production could be a fruitful approach to a better understanding of the control mechanisms.

In the current study the token-to-token variability in the production of German vowels is measured and described in the articulatory and the acoustical domains. The potential role of physical aspects, such as the contacts between tongue and palate, and the biomechanical properties of the tongue, is evaluated, in order to clarify what could be the motor control strategies underlying the production of the observed variability patterns.

2. Method

2.1. Experimental Procedure

To measure the palatal contact area and its influence on lingual vowel targets, EPG recordings (Reading EPG3) were included. Simultaneous tongue (four sensors), jaw and lip movements of three male speakers (CG, JF, and DF) of Standard German were collected by means of EMMA (AG100 by Carstens Medizinelektronik). The most anterior sensor was located around 1cm back from the tongue tip, while for the most posterior sensor the EPG palate was used as a reference. The two other sensors were placed in between, in such a way that we got similar distances between all the sensors. Sample frequencies were 100 Hz for EPG data and 200 Hz for EMMA data. The acoustical signal was recorded with a DAT and downsampled to 16kHz.

All subjects were recorded twice, once with a 5 mm thick bite block maintained between the second molars (further called BB condition) and once without bite block (further called normal condition). The BB condition was recorded in order to remove the contribution of the jaw to the token-to-token variability and to focus more

specifically on the tongue control itself. It also allowed us a comparison between experimentally measured data and data simulated by a biomechanical tongue model.

The material consisted of CVCə logatoms with either velar or bilabial stops as consonantal context and one of the 14 German vowels /iː,ɪ,yː,ʏ,eː,ɛ,øː,œ,ɑː,a,oː,ɔ,uː,ʊ/. The initial stop was voiced and the medial voiceless. All nonsense words were embedded in the carrier sentence »Sage .... bitte« ("Say .... please") and repeated 10 to 11 times.

2.2 Data Analysis

Tongue positioning for each vowel was determined as the time where most of the sensor trajectories reached a turning point during the voiced passage of the vowel, and least EPG contacts could be found.

To assess the token-to-token variability, two-sigma dispersion ellipses were computed and displayed in the plane for the four tongue sensors. The ellipses describe the spatial distribution of each sensor position at the vowel target. Variability was measured on the basis of the area of these ellipses and the angle of their main axis.

From the selected EPG patterns, a so-called Posteriority Index was computed. It corresponds to the percentage of contacts in the posterior half of the EPG palate; it gives a general information about the size of the contact area between tongue and palate. As indicated, only the back half of the EPG palate was taken into account. There are two reasons for this choice. First, we assumed that for vowel production contacts in the posterior region are more important than in the alveolar region. Second, the recordings with bite block contained a number of contacts in the front region which were not at all in agreement with basic knowledge about tongue articulation in vowel production. They were probably due to a more extensive saliva production generated by the presence of the wooden bite block in the mouth.

The acoustical signal was analyzed in the spectral domain at the same time location as the articulatory data with the software PRAAT (Boersma & Weenink, 1996). After preemphasis an LPC analysis (Burg method; 18 coefficients) was made on a 25 ms Hanning window of the signal. Just as in the articulatory domain, the variability was assessed on the basis of two-sigma dispersion ellipses in the (F1, F2) plane. Only the areas of the ellipses were considered.

3. Results

3.1 Articulatory Variability

Specific patterns of the spatial distribution of the tongue tip sensor were found compared to the other sensors. It is consistent with the fact that tongue tip can move more rapidly and more independently than the other parts of the tongue. Therefore, tongue tip is more related to the production of consonants than to the main articulation of the vowel. To study variability in vowel production, we chose (from front to back) the sensors for tongue blade (TBLADE), tongue dorsum (TDORS) and tongue back (TBACK), which describe the global positioning of the tongue body.

Looking at the areas and at the angles of orientation of the dispersion ellipses, we observed for each vowel prominent differences between the three sensors in a given condition, as well as between conditions for a given sensor. In particular, the data collected with and without bite blocks differ to a great degree. However, no systematic trend could be found across vowels which could be related to the influence of the bite block. Consequently, in the analysis of the data, the same treatment was used for the BB and the normal conditions. Areas and orientation angles of the ellipses also varied depending on vowel identity and speaker. However, a careful study of the variability patterns allowed us to find some general trends that are common to all vowels and all conditions.

Studying X-ray microbeam data in multiple repetitions of vowel [i] in different phonetic contexts, Perkell found that, for the pellets located close to the constriction location, the orientation of the main axis of the dispersion ellipses was in parallel to the outline of the vocal tract walls (Perkell & Nelson, 1985; Perkell, 1990). Our data tend to confirm these findings: it was corroborated for the three subjects under the normal condition and additionally, for two of them (CG and JD), under the BB condition. For the last case the third speaker (DF) depicted ellipses that were quite orthogonal to the palate contour for the vowels [i], [e] and [y] in both consonantal contexts.

A difference was observed in the large majority of the cases between high and low vowels. For low vowels the dispersion ellipses are essentially oriented in the high/front-to-low/back direction, similar to the orientation of the articulatory changes associated with the front raising factor found by Harshman et al. (1977). For high vowels the orientations of the ellipses are variable, because their main axes are essentially parallel to the tongue contour at the sensor location. This is true for subjects JD and CG for BB and normal conditions. Again, counterexamples were found for the third speaker in the BB condition for the vowels [i], [y], and [e], where the ellipses orientation is similar to the one of [a].

The analysis of the amount of variability, measured by the areas of the dispersion ellipses, exhibits for all speakers, all conditions and the TBACK and TDORS sensors, a smaller value for the front tense vowels [i], [e], [y] in comparison to all other vowels (for example, under normal condition and in the velar consonantal context, the ellipses areas of the TDORS sensor for speaker DF are respectively 3.7, 4.2 and 9.4 mm^2 for [i], [e] and [y], while they are 26.5, 17.2 and 15.6 mm^2 for [a], [u] and [o]). For two speakers (JD and DF) the tense vowel [u] has one of the largest variability among all the studied vowels. This is not the case for CG.

3.2 Relations between EPG and EMMA Data

In order to study the relationship between the palatal contact and the lingual vowel positions, Pearson correlation coefficients were computed between the X and Y positions of the tongue sensors and the Posteriority Index. They were systematically significant ($p<0.0001$) and showed a strong negative coefficient for the X positions and the Posteriority Index (for example, for TDORS the minimum was -0.571 and the maximum -0.897) and a strong positive correlation for the Y positions (for TDORS, the minimum was 0.809 and the maximum 0.938) for all speakers, i. e. the higher and the more anterior the tongue position, the larger is the contact area. It confirms that the articulation of front high vowels implies a noticeable amount of contacts between tongue and palate, and it suggests that the control of these articulations could make use of it, in order to stabilize the tongue at the target position.

The computation of the Pearson correlation coefficients between the ellipse areas of the three sensors and the Posteriority index tends to confirm this hypothesis. The correlations are systematically negative, and they are significant in the majority of cases for TDORS (between –0.1 and –0.79 for DF; between –0.23 and –0.82 for CG; between –0.46 and –0.91 for JD). It can be suggested that the contact between tongue and palate can explain, at least in part, the differences in variability patterns observed between high and low vowels.

4. Discussion

4.1 Articulatory Variability *versus* Acoustical Variability

Up to now our analysis of the acoustical data was limited to the tense vowels [i], [u], [e], [o], and [a] in all conditions. Concerning the amount of variability, no specific differences were observed between the normal and the BB condition. Our general observations are in agreement with classical data published in the literature: the variability is clearly larger for the low vowel [a] than for the high ones [i], [e] and [u]. It is similar to our results of the articulatory variability except for vowel [u]. Boë et al.'s (1995) study of F1 and F2 sensibilities due to changes in articulatory positions provides a framework to understand our results: [i] and [e] were found to be spectrally very sensitive to changes in tongue height, while [u] was less sensitive. Consequently, the orientation of dispersion ellipses has to be strongly controlled for [i] and [e], while the requirement is less for [u]. Perkell (1990) suggested that the strong control for the high front vowels could be facilitated by a combination of physical constraints: the lateral contacts between tongue and palate together with a stiffening of the tongue would provide a "biomechanical saturation effect", stabilizing the tongue. The negative correlation between EPG patterns and the amount of articulatory variability found in our data supports this hypothesis. However, we also showed that, in the articulatory domain, [u], [i], [e] and all high vowels share the same characteristics for the orientation of the ellipses in opposition to the orientation of the ellipses for low vowels. It suggests that some aspects of the variability patterns could be due to biomechanical factors related to anatomy of the tongue.

4.2 Variability Simulated with a Biomechanical Tongue Model

Thus, we investigated the variability due to biomechanical factors, using a 2D biomechanical tongue model (Payan & Perrier, 1997), which includes the main muscles responsible for shaping and moving the tongue in the midsagittal plane. Elastical properties of the tissues are accounted for by a finite-element method. Because of the 2D description of the model, only contacts in the middle of the palate could be taken into account. Therefore, the impact of lateral contacts cannot be studied. Following Harris' & Wolpert's (1998) hypothesis, token-to-token variability in a given phonetic environment was generated by adding white noise to the commands. Harris & Wolpert suggested adding noise all through the speech sequence. However, in the current version of the model, this approach generates dramatic computational instabilities. Hence, we generated 20 repetitions of 3 CVC sequences, where C was the velar consonant [k] and V was either [i] or [a] or [u], while changing from one repetition to the next the values of the commands at the successive targets according to a random function centered on the uncorrupted command values. All muscle commands were corrupted simultaneously, but the corrupting random function was specific for each muscle. The variance of the random function was for each muscle proportional to the square of the uncorrupted command value. Two levels of Signal-to-Noise Ratio (SNR) were tested: 10 dB and 18dB (note that according to Harris & Wolpert (1998) the SNR can be as low as 6 dB). Note also that the model is controlled according to the Equilibrium Point Hypothesis (Feldman, 1986),

which means that the commands are threshold muscle lengths, where the recruitment of α motorneurons (responsible for active forces) starts. Consequently, it should be noted that the noise was not directly added to the force commands, contrary to the suggestion made by Harris & Wolpert (1998).

For both SNR, the three vowels displayed the same kind of variability, but with different amplitudes. The spatial distributions of the nodes on the tongue surface were essentially oriented along the high/front-to-law/back direction, similar to the main factor "front raising" found by Harshman et al (1977). In addition, the generated variability was found to be larger for [i] and [u], than for [a].

These results, which are not in agreement with our experimental observations, tend to support the hypothesis that for vowel [i] the palate intervenes and permits limitation of the variability in the direction orthogonal to the palate. Thus, it modifies the variability patterns in the midsagittal plane. However, it is unlikely that such an explanation can be valid to justify the variability patterns observed for vowel [u], since the observed palatal contacts are least for this vowel. Possibly, this lack of observed contacts could be an inaccurate representation of the reality, because palatal contacts could be produced too far posteriorily and could therefore be not visible in the EPG patterns. However, another explanation could be that for the production of [u] some muscles would be more strongly controlled than others and then less corrupted by the neural noise.

5. Conclusion

The combination of the analysis of experimental data and of simulations with the biomechanical tongue model allowed us to propose some explanations of the control of token-to-token variability. For high front vowels like [i], the natural articulatory variability seems to be limited in the vertical direction due to the contacts between tongue and palate; this contributes to limit the acoustical variability. For low vowels, the observed variability seems to be essentially the natural one, due to muscle insertions and anatomy; which corresponds to quite a large acoustical variability. However, for high back vowels like [u], additional studies are necessary to understand the underlying control mechanisms of the articulatory variability.

Acknowledgments

We want to thank Dirk Fischer, Daniel Pape and Yugo Fujii, who analyzed most of the articulatory and acoustical data. Thanks are also due to Phil Hoole for comments and suggestions.

References

Boë, L.J., Badin, P. & Perrier, P. (1995). From sensitivity functions to macrovariations. *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 234-237). Stockholm, Suède, Août 1995

Boersma, P. & Weenink, D. (1996). *Praat, a System for doing Phonetics by Computer, version 3.4.* Institute of Phonetic Sciences of the University of Amsterdam, Report 132. 182 pages (also available at www.praat.org).

Browman, C.P. & Goldstein, L.M. (1990). Gestural specification using dynamically-defined articulatory structures. *J. Phonetics, 18*, 299-320.

Feldman, A.G. (1986). Once more on the Equilibrium-Point Hypothesis (λ model) for motor control. *Journal of Motor Behavior*, *18 (1),* 17-54.

Harris, C.M. (1998). On the optimal control of behaviour: a stochastic perspective. *Journal of Neuroscience Methods, 83,* 73-88.

Harris, C.M. & Wolpert, D.M. (1998). Signal dependent noise determines motor planning. *Nature, 394*, 780-784.

Harshman, R. A., Ladefoged, P. N., & Goldstein, L. (1977). Factor analysis of tongue shapes., *Journal of the Acousical Society of America, 62*, 693–707.

Lindblom, B. (1988). Phonetic Invariance and the Adaptive Nature of Speech. In *Working Models of Human Perception*. London, UK: Academic Press.

Payan, Y. & Perrier, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, *22 (2/3)*, 185-205.

Perkell, J.S. & Nelson, W.L. (1985) Variability in production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America, 77*, 1889-1895

Perkell, J.S. & Klatt, D.H. (1986). *Invariance & Variability in Speech Processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Perkell, J.S. (1990). Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 263-é88). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Stevens, K.N. & Blumstein S.E. (1981). The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale N.J.: Lawrence Erlbaum Associates.