Speech Planning for V_1CV_2 Sequences: Influence of the Planned Sequence

Pascal Perrier[1] & Liang Ma[1,2]

[1]ICP/GIPSA-lab, UMR CNRS 5216, Grenoble INP, France
[2] Laboratoire Parole et Langage, UMR CNRS 6057, Univ. de Provence, Aix-en-Provence, France

E-mail: Pascal.Perrier@gipsa-lab.inpg.fr

Abstract

The paper studies the potential influence of the structure of a language (in terms of phonological units) on the anticipatory coarticulation. It is hypothesized that speech gestures are optimally planned and that the size of the planned sequence is influenced by language constraints. The status of the syllable is studied via simulations with a speech production model and experimental data.

1 Introduction

It has been often suggested that the control of speech production sequences involves an optimal planning process in the central nervous system, which uses internal representations ([1], [2]) of the speech production apparatus ([3], [4]), in order to achieve goals in the acoustic, perceptual, and/or articulatory domain, while optimizing a speaker-oriented criterion (often called "articulatory effort"). This hypothesis is the theoretical background of the current study.

Our speech production model, called GEPETTO[1], includes a motor control model and a biomechanical model of the vocal tract ([4]). In the motor control domain the perceptual objectives of speech production are regions of the acoustic domain, which are related to targets in the space of the motor control variables. Sequence planning consists in finding optimal targets within the regions associated with the phonemes of the sequence. Once the planning has been completed, movement is achieved by shifting the motor variables at a constant rate from a target to the next and by applying them to the biomechanical model. Thus, articulatory trajectories are the result of the interaction between the motor commands at targets, the timing of their variations and the physical properties of the biomechanical model.

This paper studies for V_1CV_2 sequences the impact of the nature of the planned sequences (the whole V_1CV_2 sequence or the CV_2 syllable) on the articulatory trajectories. More specifically; we were interested in studying to which extent the phonological structure of the language could influence the kinematic characteristics of speech movements. In this context, this study focuses on the study of the potential influence of the syllable on planning. Simulations were run with three different hypotheses about the planned sequence, which account for different status of the syllable within the V_1CV_2 sequence. Simulated coarticulation patterns are compared to patterns that were experimentally observed in speakers of French and Mandarin Chinese, two syllable-based languages in which the syllable is classically considered to have different strengths in the organisation of speech ([5], [6] [7]).

2 Materials and method

2.2. Modeling

Biomechanical model

For the simulations a 2D biomechanical model of vocal tract was used, which is build around a biomechanical model of the tongue. This model includes the main muscles responsible for shaping and moving the tongue in the midsagittal plane:

[1] GEPPETO holds for "**GE**stures shaped by the **P**hysics and by a **PE**rceptually **O**riented **T**argets **O**ptimization"

posterior and anterior parts of the genioglossus (GGP and GGA), styloglossus (STY), hyoglossus (HYO), inferior and superior longitudinalis (IL and SL) and verticalis (VER). Elastic properties of tissues are accounted for by finite-element (FE). Muscles are modeled as force generators that (1) act on anatomically specified sets of nodes of the FE structure, and (2) modify the stiffness of specific elements of the model to account for muscle contractions within tongue tissues. Curves representing the contours of the lips, palate and pharynx in the midsagittal plane are added and mechanical contacts between these contours and the tongue are modeled. The jaw and the hyoid bone are represented in this plane by static rigid structures to which the tongue is attached (more details are given in [8] and [9]).

The tongue model is controlled according to the λ model [10] that specifies for each muscle a threshold length, λ, where active forces start. Tongue movements are generated from target to target by changing the λ commands at a constant rate of shift.

The whole biomechanical model is coupled with a harmonic analogue of the vocal tract, which generates formant patterns from the 3D shape of the vocal tract

Learning the relations between motor commands and acoustics: a static forward model

The internal representation used in GEPPETO's optimal sequence planning corresponds to *a static forward model* which functionally describes the relations between the λ commands and the first three formant (F1, F2, F3) patterns at targets. To learn the static forward model, 8800 different simulations were generated, which describe a large variety of tongue shapes; including those for consonants. To do so, the λ space was randomly sampled according to a uniform distribution. Formant values were computed for each simulation. Neural networks based on Radial Basis Functions (RBF) have been used to learn the relation between the λ motor commands and the (F1, F2, F3) patterns produced at target, and the mean square error on the (F1, F2, F3) patterns was less than 3% (see details in [4] and [6]).

From phonemes specification to motor commands: Influence of the planned sequence

The static forward model was used to infer from a sequence of phonemes an optimal sequence of motor commands. In this aim, for each phoneme, target regions were defined in the (F1, F2, F3) domain, and elaborated an inversion procedure to recover motor commands from the formant patterns. This inversion procedure associates the minimisation of a speaker-oriented criterion and an account for listener-oriented constraints. It aims at minimizing the global distance between the λ commands associated with the phonemes of the planned sequence while ensuring that the (F1, F2, F3) formant patterns remain within the target regions defined for these phonemes ([5], [11]).

The features of the motor control model in GEPETTO are close to those that can be found in general human motor control models. However, as compared to typical human skilled gestures, speech gestures have an additional semiotic component that is determined by the language. To study the potential influence of this language-related semiotic component onto speech movements, articulatory trajectories were generated with our model according to whether the whole V_1CV_2 sequence is planned (henceforth *sequential planning model)* or the CV_2 syllable only is planned (*syllable planning model*). In other words the above mentioned global distance is minimized either over the whole sequence or just over the CV_2 sequence.

2.2. Experimental data

To evaluate the simulated influence of the syllable strength in speech planning, simulations obtained with both planning models were qualitatively compared with experimental data collected from speakers of French and Mandarin Chinese. Speech material consists of 15 VCV nonsense words where the vowel was /a/, /i/ or /u/ and the consonant was /k/ or /t/. The words were uttered at a normal speech rate by three native French speakers and two native Chinese speakers. Each target word was embedded in a carrier sentence and each carrier sentence was repeated 10 times. The articulatory data were collected with an electromagnetic midsagittal articulograph (EMMA; AG100 Carstens Electronics). The four sensors glued on the tongue

are called T_1, T_2, T_3 and T_4, from the tongue tip to the tongue back (see [5], [6], [7] for more details).

3 Results

3.1. Measured coarticulation patterns in French and Mandarin Chinese.

Results observed for experimental data are exemplified in Figure 1.

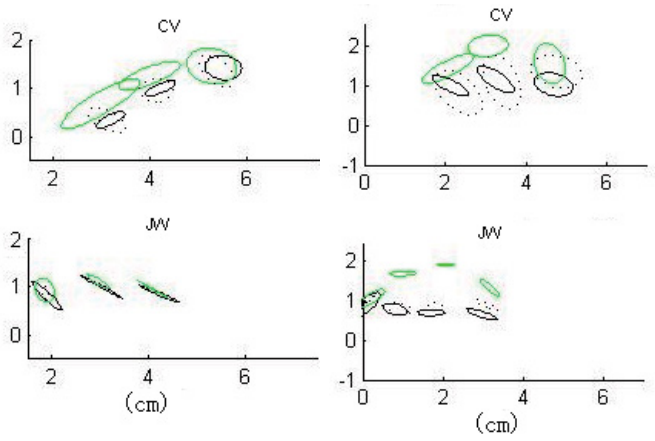

Figure 1: Variation of T_2 T_3, T_4 positions of /u/(left) and /t/ (right) in /ut/V_2 sequence,, when V_2 varies(/i/:green line; /a/:solid line; /u/:dotted line); .
Left: French speaker; Right: Chinese speaker
Top: vowel /u/; Bottom: consonant /t/

In French, V_2 influences significantly the sensor positions of both V_1 and C, and in a way that is compatible with an anticipation strategy (Sensors of get V_1 and C get closer to those of V_2. In Mandarin Chinese changes in V_2 did significantly affect sensor positions of C, but no significant consequence could be observed for V_1. In summary, in our data, anticipatory coarticulation goes back beyond the limits of the CV_2 syllable in French while it is strictly limited to the syllable in Mandarin Chinese ([5], [6], [7])..

3.2. Simulations

Figures 2 and 3 plot the positions of 4 nodes located on the upper contour of the tongue model for velar consonant /k/ and vowel /ɔ/ in /ɔkV_2/ sequences, where V_2 is either /i/, /e/, /ɛ/, /œ/, /ɔ/ or /a/. The 4 nodes are located along the tongue contour approximately at the same places than the sensors in the articulatory data presented in section 3.1. Since GEPETTO does not include a modelling of the lips, it was not possible to simulate rounded vowels in the same way as unrounded vowels. Hence, in our simulation, vowel /u/ was replaced by vowel /ɔ/ that is also articulated in the velar region of the vocal tract and does not require rounded lips to be acoustically correctly produced.
Figure 2 shows the results obtained for the *sequential planning model*, while figure 3 shows the results for *the syllable planning model.* It is important to recall here that the planning model determines the λ commands at the successive targets of the sequence, and that the actual position reached for the different elementary sounds are the results of the combined effect of these λ values, of their timing and of the dynamics of the biomechanical tongue model (mass, stiffness, damping..). For all simulations the timing of the λ commands was the same: transition time between targets was set to 40 ms, and the hold duration at targets to 100 ms.
It can be observed that sequential planning induces significant changes in tongue positioning that are consistent with anticipatory coarticulation and similar to those observed for the French speakers in the experimental study (Fig. 1). Syllable planning induces only slight changes, which are not consistent with anticipation (for example in /e/ context position is slightly backwards as compared to the /a/ context). These results are similar to those obtained for the Chinese speakers (Fig. 1). Both plannings generate variations of tongue positioning for /k/ consistent with anticipation. The variation is slightly larger though for the syllable planning. Again these results match well our experimental findings (Fig. 1).

4. Conclusion

Simulations carried out with a model of speech production accounting for optimal sequence planning and biomechanical characteristics are in good agreement with experimental observations of anticipatory coarticulation in V_1CV_2 sequences. They show that considering the length of the planning sequence in relation to the structure of the

language (here the status of the syllable) allows the replication of anticipatory coarticulation variations across languages.

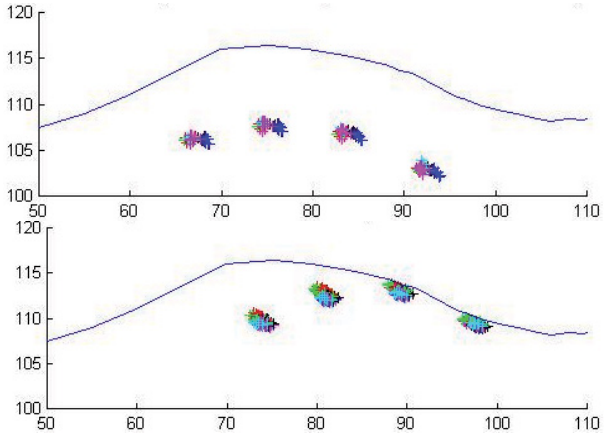

Figure 3: Node positions at target in /ɔkV$_2$/ sequences for the sequential planning model, where V$_2$ is either /i/ (green), /e (red)/, /ɛ / (magenta), /œ/ /cyan, /ɔ/ (black) or /a/ (black). The upper solid line represents the palate contour. Lips are on the left hand side. Top panel: Vowel /ɔ/; Bottom panel: Consonant /k/

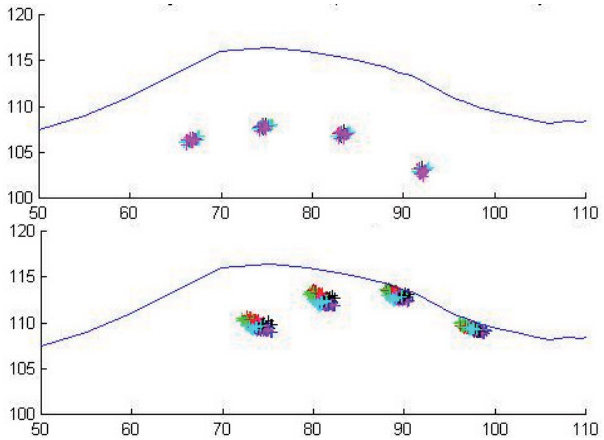

Figure 4: Node positions at target in /ɔkV$_2$/ sequences for the syllable planning model. See Figure 3 for details.

References

[1] Jordan, M.I. (1990). Motor Learning and the Degrees of Freedom Problem. In M. Jeannerod (ed.), *Attention and Performance*, Hillsdale, NJ: Erlbaum, pp. 796-836.

[2] Kawato, M., Maeda, Y., Uno, Y. & Suzuki, R. (1990) Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics, 62,* pp. 275-288.

[3] Guenther, F.H., Hampson, M. & Johnson, D. (1998) A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* pp. 611–633.

[4] Perrier, P., Ma, L. & Payan, Y. (2005) Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue. *Proceedings of Interspeech 2005*, Lisbon, Portugal, pp. 1041-1044.

[5] Ma, L., Perrier, P. & Dang, J. (2006) Anticipatory Coarticulation in Vowel-Consonant-Vowel sequences: A crosslinguistic study of French and Mandarin speakers, Proceedings of the *7th International Seminar on Speech Production*, pp.151-158, Ubatuba, Brazil

[6] Ma, L. (2008). *La coarticulation en français et en chinois : Étude expérimentale et modélisation* Unpub. Dissertation in *Cognition, Language and Education*, Université de Provence, Aix-en-Provence, France (211p.)

[7] Ma, L., Perrier, P., & Dang, J. (Submitted). Anticipatory Coarticulation in Vowel-Consonant-Vowel sequences: A crosslinguistic study of French and Mandarin Chinese. *J. Acoust. Soc. Am.*

[8] 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

[9] Perrier, P., Payan, Y., Zandipour, M. & Perkell, J. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *J. Acoust. Soc. Am., 114(3),* pp. 1582-1599.

[10] Feldman, A.G. (1986) Once more on Equilibrium Point Hypothesis for Motor Control. *Journal of Motor Behaviour, 18 (1)*, pp. 17-54.

[11] Perrier, P., Payan, Y., & Marret, R. (2004). Modéliser le physique pour comprendre le contrôle: le cas de l'anticipation en production de parole. In R. Sock & B. Vaxelaire (Eds.), *L'anticipation à l'horizon du Présent* (pp. 159-177). Pierre Margala, Sprimont, Belgium.