4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

# COMPARING TONGUE KINEMATIC AND ACOUSTIC PHASING PATTERNS FOR VOWEL QUANTITY CONTRASTS IN WOLOF

R. Sock [1] P. Perrier [2] & A. Löfqvist [3]
[1]Institut de Phonétique de Strasbourg, 22 rue Descartes - 67084 Strasbourg, France
sock@ushs.u-strasbg.fr
[2]Institut de la Communication Parlée de Grenoble - INPG, 46 ave Félix Viallet - 38031 Grenoble
perrier@icp.grenet.fr
[3]Haskins Laboratories, 270 Crown Street, New Haven-CT 06511 U.S.A.
lofquist@yalehask.bitnet

## ABSTRACT

The major question addressed in this investigation is: can one pinpoint relationships between acoustic phasing patterns and tongue body kinematic phasing patterns in the production of vocalic quantity contrasts? If so, to what extent could kinematic phasing patterns be inferred from acoustic phasing patterns? Regularities, that may unveil articulatory-acoustic relations, represent crucial information in trying to recover articulatory gestures from speech sounds. This investigation is thus, a contribution, hopefully, to understanding the complex ill-posed problem of inverse mapping.

## INTRODUCTION

A few phasing studies [1; 2; 3] have demonstrated the relevance of some temporal measures, on both the acoustic and movement levels, as domains where vowel quantity contrasts emerge distinctly. Such studies were mainly carried out on the jaw and the lips; the present investigation will attempt to extend this paradigm to the tongue body, knowing that the behaviour of this articulator is more complex (with a much higher degree of freedom) than that of the jaw. However, results obtained recently for Swedish, another quantity language, have shown significant clear-cut timings patterns not only for the lips and the jaw but also for the tongue tip and tongue body [4]. Speech rate will be varied to test for the robustness of the emerging phasing patterns, and the latter will be analyzed and discussed in terms of linguistic constraints or eventually in terms of biomechanical constraints.

## METHOD

One native speaker of Gambian Wolof — a quantity language spoken in the Gambia, Senegal and Mauratania — served as subject for this experiment. Six C1V1C2 + AM words were chosen to vary vowel quantity and quality:
/fitam/ "his courage" *vs.* /fiitam/ "his trap"
/fatam/ "its filling up" *vs.* /faatam/ "its expiration"
/futam/ "his blister" *vs.* /fuutam/ "his Fouta (Guinea)"
Each token was embedded in the interogative carrier phrase:
"lu -m ne_______ ?", meaning "What did he/she say_______?"
The speaker was instructed to produce the randomized list of utterances at a normal conversational speaking rate, and at a self selected fast speaking rate, fourteen times in each speech condition. Thus, there were 12 conditions in all: 2 speech rates x 2 vowel lengths (short and long) x 3 vowel types (i, a u).

The movement data were recorded in a sound proof anechoic chamber, using a three-coil transmitter system [5], with five transducers, the Electro Magnetic Articulograph (AG 100, *Carstens Medizinelektronik*). Receivers were placed on the vermillion borders of the upper and lower lips, the lower incisors for the jaw, and at a median position on the tongue body. In addition, a receiver placed on the upper incisors was used for correction of head movements relative to the helmet. Caution was taken during each receiver placement to insure that it was positioned at the midline with its long axis perpendicular to the sagittal plane. A simultaneous audio recording was carried out, using a helmet incorporated microphone. To obtain instantaneous velocity, the first derivative of the position signals was software calculated using a 3-point central difference algorithm. The velocity signals were smoothed using the same time constant.

Based on articulatory events in the velocity signal, two cycles were identified in the movement of each articulator (jaw, lower lip, upper lip and tongue body). These velocity cycles were determined, as the interval between successive negative or positive peaks associated with the lowering or raising movement in the production of a vowel and a consonant. An oral opening phase, associated with the production of the vowel, was defined within the oral opening cycle, and an oral closing phase, associated with the production of the consonant, was defined within the oral closing cycle. Two acoustic cycles were also defined: one as the recurrence of the onset of a clear formant structure (*i.e.* in the VC domain), corresponds to the vocalic cycle; the other as the offset of a clear formant structure (*i.e.* in the CV domain), corresponds to the consonantal cycle. Acoustic phases were specified within the appropriate acoustic cycles, as the interval that presents a stable formant structure for the vocalic phase, and as the obstruent portion for consonantal closure. It is hypothesized that oral opening and acoustic vocalic phases would reveal quantity contrasts, while oral closing and consonantal phases would highlight probable concomitant consonantal differences.

## RESULTS

Data processing is based on the percentage of time taken by each phase in its cycle. The results reported here, however, only focus on acoustic data and data for the tongue. Due to significant variability in the jaw and lip data and consequently to no coherent patterns, it is inferred that, for at least this speaker, these articulators do not play a *critical* role in the production of vocalic quantity contrasts. ANOVAs were performed on measured intervals as dependent variables and grouping factors *Quantity*, *Quality* and speaking *Rate*.

## QUANTITY CONTRASTS

### Acoustic relative timing

In the vocalic cycle, for the /i/ *vs.* /ii/ contrast (cf. Figure 1, top left panel), in normal speech, quantity contrasts emerge distinctly in this acoustic temporal domain. Patterns for short vowels are significantly different from their longer counterparts along the vocalic phase. Contrasts are in fact clear-cut, due also to highly significant cycle differences. In fast speech, the scenario is the same: quantity contrasts are maintained not only along the phase but also along the duration of the cycle. The strategies adopted to preserve the phonological contrasts under speech rate increase are as follows: a) when long vowels reduce both their cycles and the proportion of their phases under speech rate increase, their shorter counterparts do likewise, thus drifting away from a potential phonological confusion domain; b) the perceptual difference between the highly reduced long vowels in fast speech and the short vowels in normal speech, is presumably obtained by a maintained significant phase difference ($p<0.01$), as their cycle difference has been neutralized with increased speaking rate ($p=ns$).

For the /a/ *vs.* /aa/ contrast (cf. Figure 1, middle left panel), in normal speech, quantity contrasts also emerge clearly. The significant phase difference is also coupled with a mean cycle difference. In fast speech, the phase difference between the short and the long vowel is still maintained. Although both short and long vowels have their cycles compressed in fast speech, the cycle difference is still highly significant between the two entities ($p<0.01$). The strategies adopted to preserve the phonological contrasts under speech rate increase are, to some extent, similar to those used in the previous /i/ context. However, short vowels in this /a/ context are limited, for reasons related to vowel identity preservation, in the compressibility of their vocalic phases or of their overall cycle duration [6]. Thus the perceptual difference between the highly reduced long vowels in fast speech and the short vowels in normal speech, is obtained neither by a phase difference, nor by a difference in cycle duration, as phasing patterns for these two entities tend to merge ($p=ns$). It is assumed that phonological contrasts are *relayed* by other timing strategies in a different time span.

Regarding the /u/ *vs.* /uu/ contrast (cf. Figure 1, bottom left panel), phasing patterns for quantity contrast emerge clearly, in normal speech, as in the previous contexts. The distinction between the two classes is maintained in fast speech regardless of a pronounced phase and cycle compression. However, short vowels in this /u/ context do not seem to be limited in the extent of their compression, for any reasons that may be related to vowel identity preservation, as the labial consonant /f/ could contribute to enhancing the identification of this word even in an ultra-reduced vowel case. It has been reported [1] in an experiment of vowel suppression in accelerated speech for /fitam/, /fatam/ and /futam/, that it is the latter word that is identified in all cases. In other words, /f*tam/ is exclusively recognized as /futam/ when the identity of an extremely reduced vowel depends entirely on the labial context. As in the /i/ context, the perceptual difference between the highly reduced long vowels in fast speech and the short vowels in normal speech, is obtained by a maintained significant phase difference ($p<0.01$), since their cycle difference has been neutralized with increased speaking rate ($p=ns$).

### Tongue body movement relative timing

In the tongue body lowering peak velocity cycle — generally associated with the acoustic VC domain — the tongue body shows fairly coherent phasing patterns that correspond to the different phonological classes.

For the /i/ *vs.* /ii/ contrast (cf. Figure 1, top right panel), in normal speech, the distinction is based solely on the oral opening peak velocity cycle ($p<0.01$). As speech rate is increased, this distinction along the cycle is reduced; it is, nonetheless, statistically significant. The manoeuvre used to distinguish the classes is straightfoward: as the oral opening phase is not relevant in distinguishing short from long vowels in either speech rate (the relative stability observed for both classes from normal to fast speech is not in this case beneficial to the maintenance of quantity contrasts), it is the movement cycle that ensures the distinction at both speaking rates.

For the /a/ *vs.* /aa/ contrast (cf. Figure 1, middle right panel), in normal speech, the distinction is, as in the previous context, only based on a cycle difference ($p<0.01$). However, when speech rate is increased, the quantity contrast along the peak opening cycle is reinforced by an oral opening phase difference that now becomes significant. The strategies adopted to preserve the phonological contrasts under speech rate increase are the following: a) when long vowels increase the proportion of their oral opening phases under speech rate increase (*sic*) while reducing their cycles, their shorter counterparts reduce the proportion of their phases while compressing their cycles; the result is a maintained cycle difference in fast speech, coupled with an enhanced phase difference; b) long vowels in this /a/ context are limited, in the compressibility of their oral opening phases, for reasons related to linguistic coherence (avoidance of phonological class confusion) but have the possibility of increasing their phase percentage, which is finally the strategy adopted; c) consequently, pattern differences between the highly reduced long vowels in fast speech and the short vowels in normal speech, become possible through an oral opening phase difference ($p<0.01$).

For the /u/ *vs.* /uu/ contrast (cf. Figure 1, bottom right panel), distinguishing quantity contrasts is quite critical: classes are not significantly distinct at either speech rate. The linguistic contrast, observed on the acoustic level, does not therefore seem to rely on the tongue body lowering gesture. This fact calls for further investigations that should look at lip protrusion.

### Closure durations

### Acoustic relative timing

In the consonantal cycle, corresponding to the CV domain, only the *Rate* effect was significant for the consonantal phase. It should be noted that this effect reflects a tendency for consonantal phases to globally increase their proportion from normal to fast speech. However, post-hoc pairwise comparisons do not reveal any systematic coherent behaviour (in terms of direction of change). For all three vocalic contexts, consonantal patterns reduce their cycles significantly but maintain the variation of their phases within certain limits (70% and 75% of the cycle) regardless of increase in speech rate. If such a behaviour is far from portraying any sort of relative invariance, it depicts a constrained behaviour, presumably linked to both linguistic and biomechanical constraints.

## Movement relative timing

In the tongue body raising peak velocity cycle — generally associated with the acoustic CV domain — and for all of the three vocalic contexts, oral closing patterns show a significant reduction of their cycles but a constrained variability of their oral closing phases within certain limits, regardless of increase in speech rate (between 50% and 60% of the cycle).

## SUMMARY

These results corroborate major timing results obtained for quantity contrasts using this paradigm.

On the acoustic level: a) the vocalic cycle is an efficient temporal domain where quantity contrasts are exemplified. Such results are in line with reported data [1; 2]. The vocalic phasing patterns corresponding to the different linguistic entities emerge clearly in this temporal span in normal speech and are more or less maintained in fast speech, thus confirming the relevance of such measurements in speech timing studies and quantity contrasts; b) consonantal phasing patterns in the consonantal cycle do not reveal concomitant consonantal differences. In other words, the different linguistic categories (short *vs.* long vowels) show comparable consonantal patterns, across rate changes, seemingly indicating some constraints in consonantal phase variability. In a similar study [4] it has been shown that this cycle is efficient in portraying not only consonantal quantity contrasts but also vocalic contrasts, where systematically, phonologically short vowels are followed by long consonants and phonologically long vowels by short consonants. This is not the syllablic structure observed for Gambian Wolof according to data available in the literature.

On the movement level: a) the peak velocity cycle of the tongue body's vertical displacement is a relevant temporal span to observe for vocalic quantity contrasts. The oral opening phasing patterns corresponding to the different linguistic entities emerge clearly in this temporal span in normal speech and in some instances are maintained in fast speech. This movement cycle also proved efficient for the jaw in its vertical displacements to produce vocalic quantity contrasts [2; 3]; b) oral closing phasing patterns in the oral closing cycle do not reflect a complementary distribution of the quantities [7]. This means that the different linguistic categories (short *vs.* long vowels) show comparable consonantal patterns. Again, if this domain is not useful in distinguishing linguistic classes, it may be pertinent in unveiling relatively constrained behaviours [8].

## Comparing levels

These results also show that, if movement patterns are quantitatively different from acoustic ones, regarding their cycles and consonantal phases, a comparable timing relationship seems, however, to be the phase proportion taken by the acoustic vocalic and the tongue movement opening phases in the acoustic and velocity cycles respectively: the values for these phases are comparable, in both normal and fast speech for short and long vowels. Moreover, it is in these domains, where quantity contrasts are better portrayed, that acoustic and movement phasing patterns are *structurally* similar. Although consonantal phasing patterns in the acoustic domain are quantitatively different from those in the movement domain, they too are structurally similar and also call for more qualitative pattern analyses to be able to push conclusions further.

After having thus identified the appropriate temporal domains for such an investigation, it should be possible to infer kinematic phasing patterns from acoustic patterns, at least, from a global structural perspective. These results confirm our initial hypotheses and also corroborate previous findings obtained for the tongue body in related studies. More data, together with more rigorous and adequate statistical analyses are needed in this area to be able, not only to quantify differences between acoustic and kinematic phasing patterns, but also to establish systematic relations between these patterns.

## ACKNOWLEDGMENTS

Our sincere thanks go to Philippe Mahé for the experimental setup and data analyses. This research was supported by an ESPRIT-BR Project 6975—Speech Maps, NIH Grant DC-00865, from the National Institute of Deafness and Other Communication Disorders and the Fyssen Foundation.

## REFERENCES

[1] Sock R. (1984) Une compensation temporelle, en fonction de la vitesse d'élocution dans le timing de l'opposition de quantité vocalique du wolof de Gambie. *Bulletin de l'Institut de Phonétique de Grenoble* 13, 25-84.

[2] Delattre C. Jomaa M. Al Dossari A. Worley C. Sock R. (1990) Comparaison articulatori-acoustique des structures temporelles en arabe et en français. Où peut-on séparer les classes dans les VC ? *18èmes Journées d'Etudes de la Parole du GCP de la SFA*, 113-118.

[3] Boussaffa F. Jomaa M. Sock R. (1991) Les contraintes temporelles des types consonantiques sur le timing mandibulaire de la quantité en arabe tunisien. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 2, 306-309.

[4] Sock R. Löfqvist A. (1995) Quantity contrast in Sweedish. Kinematic and acoustic patterns. *Proceedings of the 13th International Congress of Phonetic Sciences.* Stockholm, August 13-19.

[5] Perkell J.S. Cohen M.H. Svirsky M.A. Matthies M.L. Garabieta I. Jackson M.T. (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America 92, 3078-3096.*

[6] Abry C. Orliaguet J.-P. Sock R. (1990) Patterns of speech phasing. Their robustness in the production of a timed linguistic task: single versus double (abutted) consonants in French. *CPC, European Bulletin of Cognitive Psychology 10, 269-288.*

[7] Engstrand O. Krull D. (1994) Durational correlates of quantity in Sweedish, Finnish and Estonian: cross-language evidence for a theory of adaptive dispersion. *Phonetica 51, 80-91.*

[8] Delattre C. Perrier P. (1991) Phase modifications in tongue movements across speech rate variations: influence of consonantal articulation. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 3, 26-29.

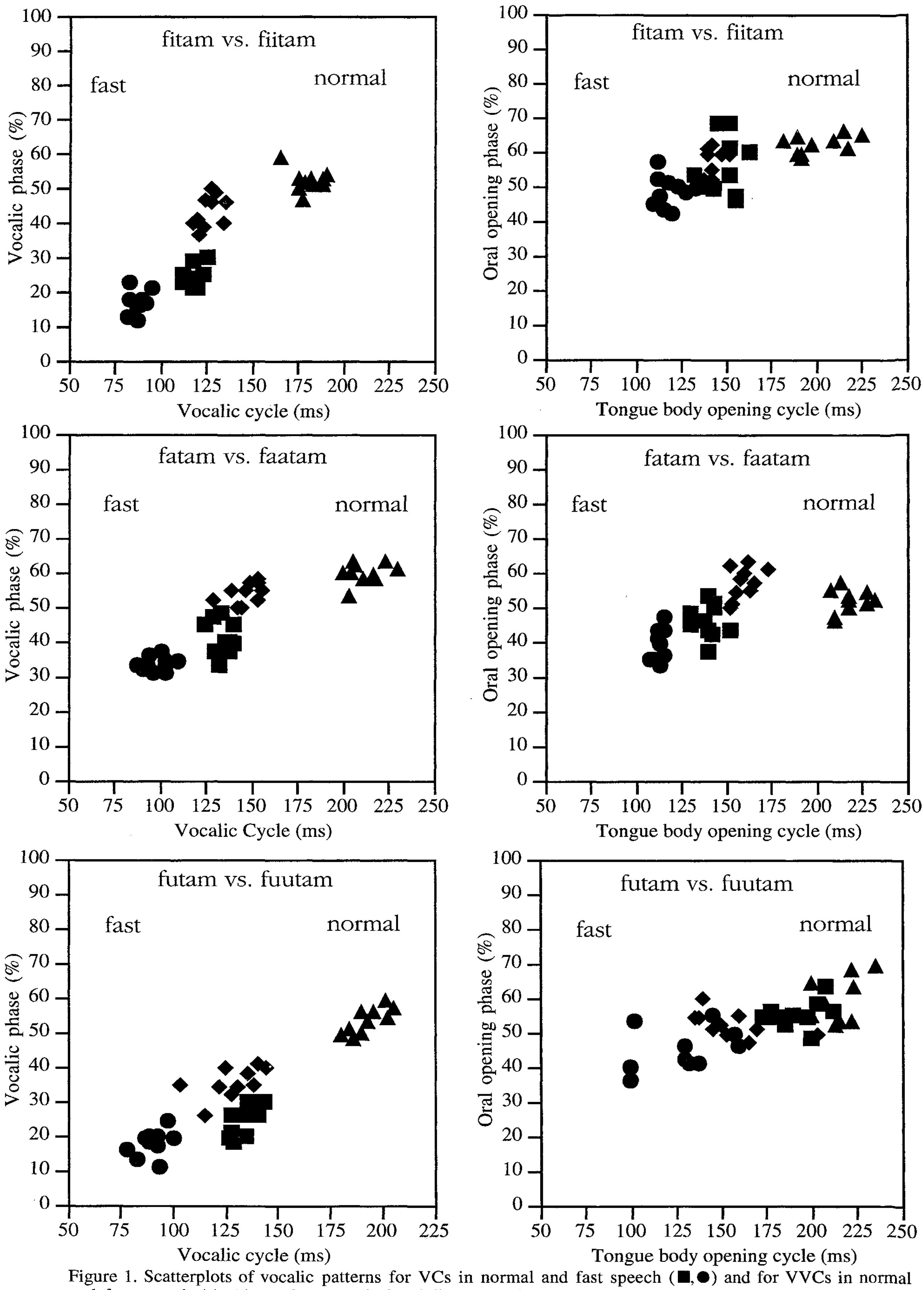


Figure 1. Scatterplots of vocalic patterns for VCs in normal and fast speech (■,●) and for VVCs in normal and fast speech (▲,◆) on the acoustic level (left panels) and on the tongue body movement level (rignt panels) for Wolof words. See text for details.