



HAL
open science

Gestural cost of articulatory gestures: insight from segmental durations

Martine Toda

► **To cite this version:**

Martine Toda. Gestural cost of articulatory gestures: insight from segmental durations. Labphon 12, 2010, Albuquerque, United States. halshs-00677144

HAL Id: halshs-00677144

<https://shs.hal.science/halshs-00677144>

Submitted on 12 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Auteur

Martine Toda
LPP, UMR 7018 CNRS/Université Paris 3 et LTCI, UMR 5141 CNRS/Telecom
ParisTech

Title

Gestural cost of articulatory gestures: insights from segmental durations

Abstract

Focusing on articulatory gestures as the minimal units of the time-varying behavior of speech organs, this paper discusses, based on a large corpus, how segmental durations are suggestive of the complexity or magnitude of articulatory gestures and their organization in connected speech.

A body of evidences tends to show that segmental duration is not random. For illustration, the MOCHA database (Wrench, 1999) contains 13960 phoneme utterances (durations of vowels and consonants, based on the labeling data provided), which ranges from 20 up to about 350 ms. Interestingly, there exist a significant correlation of segmental duration between two subjects reading the same corpus ($r=0.66$; $p<0.001$). The phoneme utterances produced short by the female subject 'fsew' are also short for the male subject 'msak', and so on (figure 1a). This indicates that segmental duration is primarily conditioned by linguistic factors, rather than being due to individual or random variation.

While segmental durations are well related to syllable structure (e.g. Maddieson, 1985), the intrinsic durations of the various phoneme categories (e.g. in Klatt, 1979) are not accounted for by any theory. A basic explanation would propose that segmental durations depend, at least in part, on the articulatory complexity or distance required to move from a target to the next. This proposal, in line with Redford (2004), is consistent with the fact that consonants are shorter when produced within consonant clusters (whose articulatory targets are closer) than when produced in intervocalic position (e.g. Kühnert et al, 2006).

In this study, we will focus on /s/, a sibilant fricative, that is known to be resistant to coarticulation, and to require a precise tongue posture in order to be correctly produced (e.g. Stone et al., 2007). /s/'s duration is therefore expected to be more dependent on its phonetic context rather than on its own target undershoot (under the effect of prosodic factors or whatsoever). We hypothesize that longer utterances of /s/ (see Figure 1b) would be explained by a larger distance to the preceding and following targets, or the involvement of a larger number of articulators (e.g. tongue tip *and* tongue dorsum).

In order to verify this hypothesis, the gestural cost cued by acoustic segmental duration will be compared against the gestural cost cued by articulatory parameters (distance between articulatory targets and criticality of the articulators - cf. Ananthakrishnan and

Engwall, 2008).

By quantitatively assessing the cost of articulatory gestures, we hope to open new perspectives on the understanding of the various syllable structures in the world's languages, and their relation to phoneme inventories.

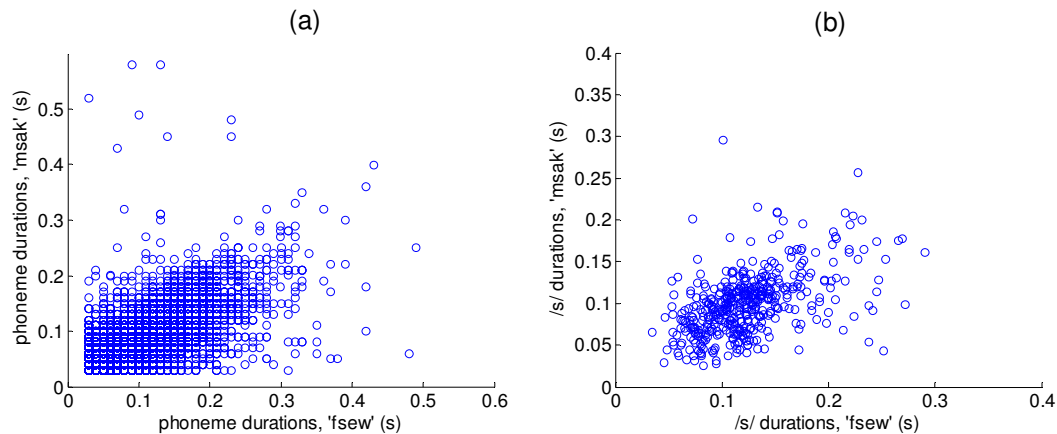


Figure 1. (a) Segmental duration of 13960 phoneme utterances, produced by two subject of the MOCHA database : 'fsew' in x axis and 'msak' in y axis. (b) Segmental duration of 437 utterances of /s/, whose boundaries were corrected (for a better time precision) by means of an automatic alignment algorithm that uses changes in magnitude of the signal envelope and zero-crossing rate.

References

- Ananthakrishnan, Gopal and Olov Engwall, 2008. Important regions in the articulator trajectory, In *Proc. 8th ISSP*, Strasbourg (France), p. 305-308.
- Klatt, Dennis H., 1979. Synthesis by rule of segmental durations in English sentences. In Lindblom and Öhman, *Frontiers of speech communication research*, Academic Press: 287-300.
- Kühnert, Barbara, Phil Hoole, and Christine Mooshammer, 2006. Gestural overlap and C-center in selected French consonant clusters. In *Proc. 7th ISSP* 327-334.
- Maddieson, Ian, 1985. Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic Linguistics. Essays in Honor of Peter Ladefoged*, 203-221. New York: Academic Press.
- Redford, Melissa A. 2004. Origin of consonant duration patterns. In *Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele et al., 54-61. Somerville, MA: Cascadilla Proceedings Project.
- Stone, M., G. Stock, K. Bunin, K. Kumar, M. Epstein, C. Kambhamettu, M. Li, V. Parthasarathy & J. Prince, 2007. Comparison of speech production in upright and supine position. In *JASA* 122 (1), 532-541.
- Wrench, Alan, 1999. The MOCHA-TIMIT darticulatory database.
<http://www.cstr.ed.ac.uk/research/projects/artic/mocha.htm>