

# Covariations Of English Segmental Durations Across Speakers

Jiahong Yuan

University of Pennsylvania, U.S.A.

jiahong@ling.upenn.edu

## Abstract

This study investigated the ways in which segmental durations co-vary across speakers in a large speech corpus. We found that pauses were negatively correlated with other phones. Speakers who produced longer speech phones tended to have shorter pauses. The approximants had little correlation with the other phone classes. This suggests that the durations of the approximants are controlled by speaker-dependent mechanisms.

**Index Terms:** segmental duration, covariation, timing control

## 1. Introduction

Durational characteristics of segments have been widely studied, especially in terms of the linguistic or paralinguistic factors that affect segmental durations [1]. This study examined the durational patterns from a different perspective. We were interested in finding out how the durations of the segments co-vary across individual speakers; that is, whether the durational difference between two segments is consistent across speakers or shows speaker characteristics. The results will provide insight into the mechanisms of timing control in speech production.

## 2. Data, Methods, and Results

The SCOTUS corpus includes more than 50 years of oral arguments from the Supreme Court of the United States. We extracted and utilized the “clean” turns (based on the transcripts) of eight Justices from 78 hour-long recordings from 2001. The phone boundaries were automatically determined, using word pronunciations from the CMU pronouncing dictionary and a forced aligner trained on the same data with the HTK toolkit. We tested the same forced aligner on the TIMIT corpus, where the average difference between the forced aligned phone boundaries and the manually labeled phone boundaries in TIMIT was about 15 milliseconds.

Our dataset contained 758,010 phones from eight speakers. The durations of the phones were calculated from the boundaries in the forced alignment. We calculated the mean durations of nine broad phone classes: Affricates (‘AF’); Approximants (‘AP’); Fricatives (‘FR’); Nasals (‘NS’); Stops (‘ST’); Reduced vowels (‘V0’); Primary-stress vowels (‘V1’); Secondary-stress vowels (‘V2’); and Pauses (‘SP’). Each phone class had eight means (one from each speaker). These mean numbers were used in the correlation analysis between different phone classes.

Figure 1 shows that pauses have negative correlations with other speech phones. Speakers who had longer speech phones tended to have shorter pauses. The histogram of the pause duration shown in Figure 2 suggests that there are more than one type of pauses; this conclusion is consistent with the results of previous studies [2]. The correlations between speech phones and shorter pauses (< 500 milliseconds) and between speech phones and longer pauses (>= 500

milliseconds) were less significant, but both were still negative.

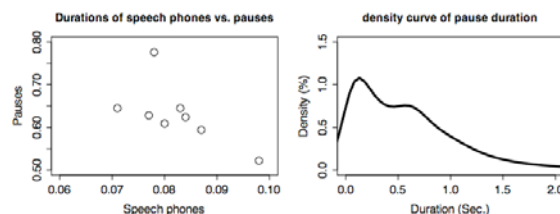


Figure 1. Durations of eight speakers. Figure 2. Pause durations.

Excluding pauses, we calculated the correlations of the mean durations between every pair of the other phones. The Pearson’s  $r$  coefficients are listed in Table 1. The primary-stress vowels and secondary-stress vowels had the highest correlation ( $p < 0.001$ ). The lowest correlations were between the approximants ( $l, r, w, j$ ) and the nasals, stops, and reduced vowels ( $p > 0.2$ ).

Table 1. Correlation coefficients between phone classes.

	AF	AP	FR	NS	ST	VO	V1	V2
AF	1.0	0.75	0.91	0.67	0.76	0.82	0.67	0.56
AP	0.75	1.0	0.86	0.46	0.47	0.46	0.60	0.53
FR	0.91	0.86	1.0	0.68	0.75	0.69	0.71	0.57
NS	0.67	0.46	0.68	1.0	0.95	0.88	0.95	0.92
ST	0.76	0.47	0.75	0.95	1.0	0.95	0.89	0.80
VO	0.82	0.46	0.69	0.88	0.95	1.0	0.83	0.77
V1	0.67	0.60	0.71	0.95	0.89	0.83	1.0	0.98
V2	0.56	0.53	0.57	0.92	0.80	0.77	0.98	1.0

Figure 3 illustrates the Euclidean distances of the phone classes from a multidimensional scaling analysis; the distance between two phone classes was defined as one minus their correlation coefficient. Figure 3 depicts four groups of phones: primary-stressed vowels, secondary-stressed vowels, and nasals; stops and reduced vowels; fricatives and affricates; and approximants. The phone classes had higher correlations with each other within each group and probably share the same mechanism of timing control for speakers. Between-group differences may suggest different timing control mechanisms; for example, the approximants had little correlation with the other phones. Their durations are unlikely to be controlled by speaker-general mechanisms such as speaking rate, intrinsic duration, and linguistic functions of duration.

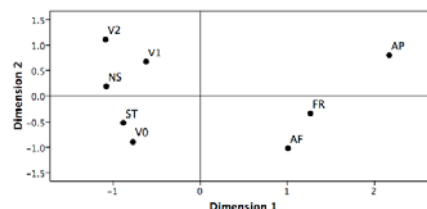


Figure 3. Euclidean distance model.

[1] Klatt, D. H. 1976. “Linguistic uses of segmental duration of English: acoustic and perceptual evidence”, *Journal of the Acoustical Society of America*, 59: 1208-1221.

[2] Campione E. and Véronis J., “A Large-Scale Multilingual Study of Silent Pause Duration”, *Proc. of Speech Prosody*, 2002.