

DYSARTHRIC SPEAKERS' INTRINSIC VOWEL DURATIONS

Stefan Werner^a & Einar Meister^b

^aUniversity of Eastern Finland, Finland; ^bTallinn University of Technology, Estonia

stefan.werner@uef.fi; einar@ioc.ee

ABSTRACT

This study uses the Nemours Database of Dysarthric Speech and the Buckeye Corpus of Conversational Speech to look into differences in the way vowel quality correlates with intrinsic duration in typical and non-typical populations. Results based on speech material from ten dysarthric subjects indicate that intrinsic vowel duration may indeed play a role as a parameter for acoustic classification.

Keywords: dysarthria, intrinsic duration, vowel

1. INTRODUCTION

Typical acoustic assessment measures for dysarthric speech include, among others, f_0 and formant variation, jitter, shimmer and HNR, but seldom intrinsic duration of speech segments. The aim of this paper is to explore whether anomalies in the intrinsic duration of monophthong vowels occur systematically in clinical populations and thus also could be instrumentalized as assessment criteria.

Differences in typical vowel duration depending on vowel quality, in particular vowel height, have been found in various studies on a number of languages (for a comprehensive overview of classical work in this area, see [1], more recent research includes e.g. [5, 6]). It is generally assumed as a universal trend that there is some correlation of degree of vowel openness with so-called intrinsic duration. Detailed attested duration rank orders vary from language to language but a low vowel like /a/ always seems to exhibit longer average durations than high vowels like /i/ and /u/.

[2] has reported overall lengthening of vowels, but preservation of typical intrinsic duration contrasts, in French subjects with Parkinsonian dysarthria. For our study we compare the vowel duration data of an American English dysarthric speech database with the data from a large database of American English conversational speech. We will focus particularly on the relationship between durations of low and high vowels.

2. DATA AND METHOD

One of the few available corpora of annotated untypical speech is the Nemours Database of Dysarthric Speech [7]. It contains data from eleven male speakers and the short test sentences produced by ten of them are labeled at the phoneme level (the eleventh speaker's intelligibility was too low for phonemic annotation). The corpus also provides background information on the subjects, including their score on the Frenchay dysarthria assessment [3]. From the pooled label files we extracted timing information for all annotated vowels, a total of 5,130.

As baseline comparison we used the Buckeye Corpus of Conversational Speech [8] which contains data from forty female and male speakers, also labeled at the phoneme level (utilizing the same TIMIT [4] annotation scheme as the Nemours database). After converting the xwaves format label files containing only segment endpoints to an interval format we again pooled all annotations and extracted the vowel timings, adding up to a sample size of 330,942.

The vowel labels with their beginning and end times were read into R [9] for calculation of the vowel durations and subsequent data analysis and visualization. We will not present any significance test results (e.g. t-test) for typical vs. untypical duration means or similar, since due to the large sample size virtually any group comparison will turn out highly significant.

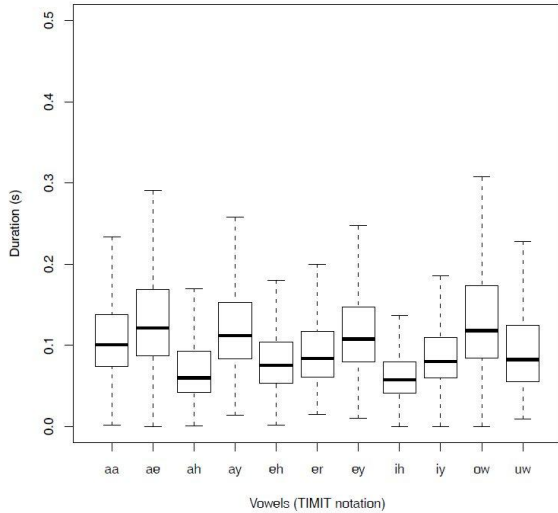
3. RESULTS

3.1. Overall distribution

Our baseline data from the Buckeye corpus gives a clear picture of the differences in intrinsic duration. Despite the considerable variation in vowel durations, not only due to the large number of speakers but even more so to the conversational setting of the interviews in which the speech data was elicited, there seems to be a systematic distribution of segment durations based on vowel

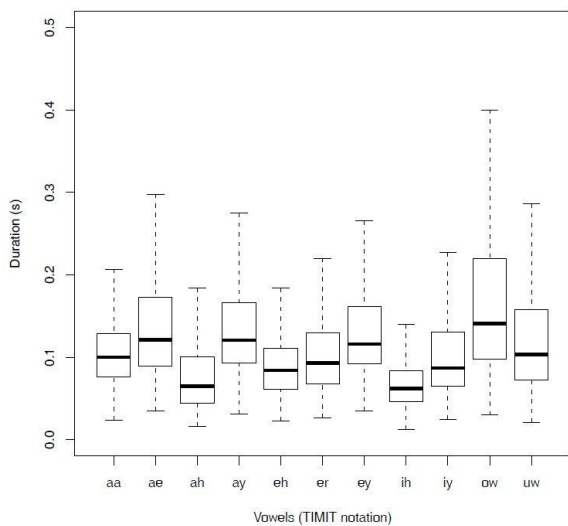
quality, as can be seen in the boxplots for all vowel types (Fig. 1).

Figure 1: Boxplots of all vowel durations in the Buckeye corpus. (Outliers omitted.)



The same overall picture also emerges from individual speakers' data, as can be seen in Fig. 2 which presents as an example speaker S04's duration data from 7,420 vowels: the ranking order of duration median values is identical.

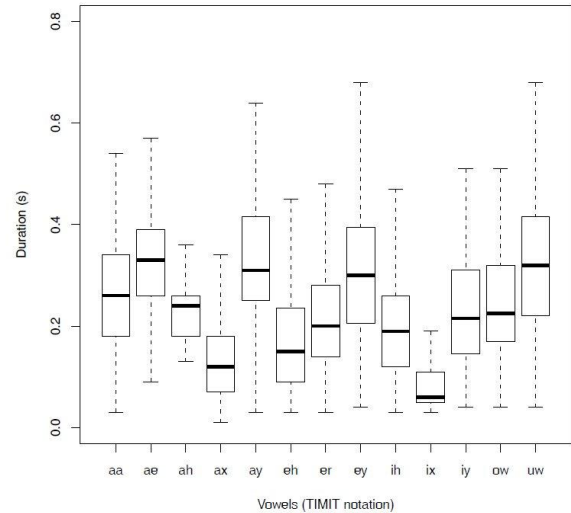
Figure 2: Boxplots of Buckeye speaker S04's vowel durations. (Outliers omitted.)



A first comparison with the Nemours data (Fig. 3) immediately shows three striking differences:

- all durations tend to be higher,
- the spread of the duration data is considerably larger and
- the ranking order of vowel types' durations is not completely the same.

Figure 3: Boxplots of all vowel durations in the Nemours corpus. (Outliers omitted.) In addition to the Buckeye vowel labels Nemours also makes use of ax for the schwa vowel (coded as iy in the Buckeye corpus) and ix for centralized ih.



In the following we will concentrate on the subset of monophthongs whose intrinsic durations have received most attention in the literature.

3.2. Low vs. high monophthongs

Figs. 4 and 5 show, plotted to the same scale, the duration distributions for /a/, /æ/, /i/ and /u/ in the Buckeye and Nemours corpora, respectively. The duration relations between most of the vowel types in the Nemours data seem to follow the expected ranking, but /u/ (uw) makes for an obvious exception: it tends to be almost as long as the longest (low) monophthong in the dysarthria dataset, /æ/ (ae).

Figure 4: /a/, /æ/, /i/ and /u/ in the Buckeye corpus. (Outliers omitted.)

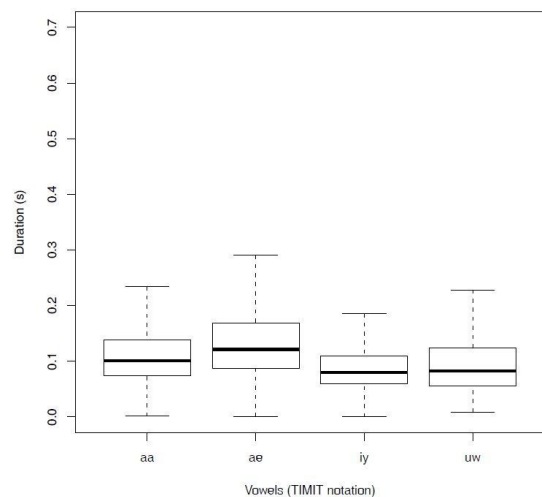
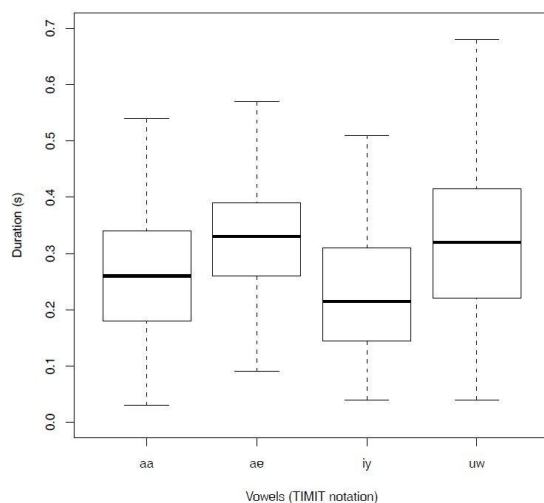


Figure 5: /a/, /æ/, /i/ and /u/ in the Nemours corpus. (Outliers omitted.)



The high number of outliers with high duration values slightly changes the picture when, in addition to medians, also mean values are compared, as in Table 1; but also in terms of the arithmetic mean there remains a discrepancy between the status of /u/ in the Buckeye corpus and the dysarthria data.

Table 1: Median and arithmetic mean values for vowel durations (ms) in the Buckeye and Nemours corpora.

	Buckeye		Nemours	
	Median	Mean	Median	Mean
/a/	101	114	260	268
/æ/	122	137	330	336
/i/	80	94	215	240
/u/	83	102	320	328

4. DISCUSSION

On the basis of our comparative analysis of vowel duration distributions in typical and clinical populations so far, it seems that intrinsic duration of vowels is a cue to articulatory symptoms that occur in dysarthria. At the same time, the limited size of the Nemours corpus and the missing opportunity to include also speech from other untypical populations in the study make it important to abstain from too far-reaching conclusions: the longer durations (caused at least partly by overall slower articulation rates) and their wider dispersion in the Nemours data are most likely not exclusively specific to dysarthria, nor can we know whether the surprisingly long /u/ vowels are also a feature of larger populations, or then perhaps something characteristic for the

particular selection of dysarthria manifestations present in the ten subjects under investigation.

Another limitation of our study is that we ignored systematic influences on vowel duration from syllable prominence and changes in articulation rates. A more subtle analysis of the duration data taking into account these additional factors should yield more valid results than our first exploration presented here.

Nevertheless our results indicate that further clinical research into intrinsic duration is worthwhile – and, en passant, also provide additional evidence for earlier hypotheses about vowel height-related duration differences in typical populations.

5. ACKNOWLEDGEMENTS

The authors are grateful to an anonymous reviewer for pointing out deficiencies in the originally submitted text.

6. REFERENCES

- [1] di Cristo, A. 1978. *De la Microprosodie à l'intonosyntaxe*. Aix-en-Provence: Université de Provence.
- [2] Duez, D. 2009. Segmental duration in Parkinsonian French speech. *Folia Phoniatrica et Logopaedica* 61(4), 239-246.
- [3] Enderby, P.M. 1983. *Frenchay Dysarthria Assessment*. London: College Hill Press.
- [4] Garofolo, J.S., et al. 1993. TIMIT acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium Philadelphia*.
- [5] Kouznetsov, V. 2001. Perceptual role of inherent vowel duration as distinctive feature in Russian. *Proceedings of the XI Session of the Russian Acoustical Society Moscow*, 443-447.
- [6] Meister, E., Werner, S. 2006. Intrinsic microprosodic variations in Estonian and Finnish: Acoustic analysis. *Fonetiikan Päevä 2006 (The Phonetics Symposium 2006)* Helsinki, 103-112.
- [7] Menendez-Pidal, X., Polikoff, J.B., Peters, S.M., Leonzo, J.E., Bunnell, H. 1996. The nemours database of dysarthric speech. *Proceedings of the Fourth International Conference on Spoken Language Processing Philadelphia PA, USA*.
- [8] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., Fosler-Lussier, E. 2007. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University. <http://www.buckeyecorpus.osu.edu>
- [9] R Development Core Team. 2011. R: A language and environment for statistical computing. <http://www.R-project.org>