ANALYSIS OF DYSARTHRIC SPEECH BY MEANS OF FORMANT-TO-AREA MAPPING

S. Ciocea (1)*, J. Schoentgen (1)†, L. Crevier-Buchman (2)

(1) *Laboratory of Experimental Phonetics, Institute of Modern Languages and Phonetics,*
CP110, Université Libre de Bruxelles, Av. F. D. Roosevelt, 50, B-1050 Brussels, Belgium.
Tel. +32 2 650 2010, Fax: 32 2 650 2007, E-mail: sciocea@ulb.ac.be
(2) *Laboratoire Voix, Biomatériaux et Cancérologie ORL, Service d'ORL et de Chirurgie de la Face et du Cou, Hôpital Laënnec, Paris, France*

ABSTRACT

This article presents a preliminary study of dysarthric speech by means of formant-to-area mapping. Dysarthria is a speech impairment which is the result of paralysis or ataxia of the speech muscles. Formant-to-area mapping is the inference of the shape of a tract model via observed formant frequencies. The corpus is composed of vowel-vowel sequences [iaia] produced by speakers suffering from amyotrophic lateral sclerosis (ALS) and normal speakers. The results show that the shapes and movements of the acoustically mapped area function models are typical of the motions and postures of the vocal tracts of ALS speakers.

1. INTRODUCTION

The purpose of our study is to examine whether the kinematics of an equivalent area function model would be appropriate for characterizing speech disorders. Here, an equivalent area function model designates an area function with the same three first eigenmodes as a speaker's vocal tract. Given the non-uniqueness of the formant-to-area transformation, the eigenmode equivalence does not guarantee the anatomical accuracy of the vocal tract shape inferred from the formant frequencies observed.

The reasons for which an acoustic-to-geometric transformation might be deemed useful in the framework of the study of dysarthric speech are the following. (i) The tract shape related features obtained by means of formant-to-area mapping are acquired non-invasively without special purpose hardware (but with no guarantee of perfect anatomical accuracy). (ii) Due to the aberrant control of the vocal tract, speech disorders bring about a greater variability of the formant movement patterns than normal speech. This constitutes an opportunity for assessing acoustic theories and analysis methods developed for speech produced by normal subjects [10]. (iii) Phonetic descriptions of speech are anyway based on schematic representations of the vocal tract, omitting many anatomical details. Some phonetic attributes such as zones of articulation and timing can be recovered by means of formant-to-area mapping. (iv) Representations of the speech signal in terms of equivalent tract shapes are easier to interpret than formant trajectories. The reason is that complicated formant movements may be the outcome of several (quasi)simultaneous simple opening/closing gestures occurring at different places within the vocal tract. (v) Anatomical data obtained by direct observation can constrain formant-to-area mapping and vice-versa. Indeed, formant-to-area mapping can be used to post-synchronize acoustic and anatomic data recorded asynchronously [8]. (vi) Formant-to-area mapping may suggest the relevance of features related to timing and tract shapes that could be explored further via direct observation.

The presentation introduces an analysis, by means of formant-to-area mapping, of reiterated vowel-vowel sequences produced by dysarthric speakers. Dysarthria is a speech impairment which is caused by the paralysis or ataxia of speech muscles. There are three main groups of dysarthria: cerebelar, extrapyramidal and paralytic [1]. Amyotrophic lateral sclerosis (ALS) is a paralytic dysarthria characterized by the following acoustic symptoms [2]. (1) The speech rate is slower than normal. (2) The targets of the formant movements are not well defined. (3) Nasality is excessive. (4) The distinction between vowels and consonants is minimal, with a tendency towards a single type of vowel or consonant.

2. FORMANT-TO-AREA MAPPING

Formant-to-area mapping is the inference of the shape of a vocal tract model via observed formant frequencies. Here, it consists of the direct calculation

---

*Action de Recherche Concertée, Communauté Française de Belgique

†National Fund for Scientific Research, Belgium

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

of the time derivatives of the cross-sections and the length of a vocal tract model so that the time derivatives of the observed formant frequencies and the model's eigenfrequencies match [6; 5; 7]. The vocal tract model is a concatenation of cylindrical tubelets. Time derivatives of the tubelet cross-sections are obtained by solving a linear algebraic system of equations. The derivatives are then numerically integrated to arrive at the cross-section movements. Since more than one area function is compatible with the observed formant frequencies, constraints are applied to the area function movement to select a unique solution.

3. CORPUS AND METHODS

The corpus was composed of vowel-vowel sequences [iaia] produced by 8 speakers (2 males and 6 females) suffering from amyotrophic lateral sclerosis and 5 normal speakers (2 males and 3 females).

The formant movements were automatically extracted by means of linear predictive coding and cepstral analysis. Since formant trajectories obtained from dysarthric speakers may be very noisy, the formant frequencies were manually estimated and automatically refined by means of the *SNORRI* speech analysis software, as required. When several plausible candidates for the same formant were obtained within the same analysis window, all possible combinations of the first three formants were selected and, together with estimated glottal frequency, inputted into Klatt's formant synthesizer [3]. The synthetic stimuli arrived at were then perceptually evaluated and the formant triplet giving rise to the synthetic stimuli nearest to the natural one was selected for formant-to-area mapping.

The area function model consisted of a concatenation of 8 cylindrical tubelets that had lengths equal to $L/10$, $L/15$, $2L/15$, $L/5$, $L/5$, $2L/15$, $L/15$, $L/10$ respectively, where L is the distance between the glottis and the lips [4]. This choice was the outcome of a previous study showing that this model fitted measured area functions geometrically better than the corresponding model consisting of 8 tubelets of equal lengths. The cross-section area of the tubelet adjacent to the glottis was fixed at 2 cm^2 [6; 5]. The areas of the other seven cross-sections were variable, and the total length, L, depended via formula (1) on the cross-section area of lip tubelet A_8. The symbol cm means that length L was measured in centimeters and cross-section A_8 in square centimeters. Consequently, the area function model had 7 independent control parameters.

$$\frac{L}{cm} = 20 - 0.5\frac{A_8}{cm^2}. \tag{1}$$

Formula (1) was obtained by means of the regression analyses of published data [9].

Radiation at the lips and wall vibration losses were taken into account in the wave propagation model.

Formant-to-area mapping was constrained by keeping as small as possible the Euclidean distance between the acoustically mapped and uniform area function models.

To enable speaker comparisons, two vowel-vowel cycles [aiaia] typical for each speaker were selected and aligned in time with reference to mid-vowel [a].

The formant frequencies in Table 1 were inputted into the formant-to-area map.

4. RESULTS AND DISCUTION

In Table 1 the formant frequency movements are plotted separately for normal and dysarthric subjects. Visual inspection of the formant frequency movements of the VV-cycles of the normal speakers shows that they have similar amplitude and timing patterns. On the contrary, the formant frequency movements of dysarthric speakers are noisier and more variable in time and amplitude. Moreover, they are slower. However, two classes of dysarthric speakers can be detected. The formant frequency movements corresponding to medium-ALS speakers are more regular. The irregular patterns are those of the sever ALS speakers.

The cross-section movements inferred by means of formant-to-area mapping are shown in Figure 3. Given the phonetic quality of the vowels involved, the relevant area function zones are A_3 and A_4 for [a], and A_5 and A_6 for [i]. The slow, noisy and ill-targeted movements of the formants transform into similarly patterned cross-section motions. The distinction between medium and severe ALS speakers is maintained. The modeled area function zones with smaller than normal movement amplitudes agreed with the zones of the speaker's vocal tract, but with mobility deficits observed by means of video fiberscopy. Table 2 shows the trajectory of cross-sections A_3 and A_6 of a normal and a dysarthric speaker with low cross-section A_3 mobility, which stayed near 6 cm^2. The residual horizontal motion is related to excess noise in the formant movements.

However, the speech signal was easier to segment via the computed cross-sections than by means of the formant movements observed. Indeed, it was enough to track the smallest cross-sections which are, acoustically and phonetically speaking, the most important. For example, to locate the center of vowel [a] in an [iai] context it was enough to locate the cross-section minimum in zones 3 or 4.

To conclude, it is suggested that formant-to-area mapping could be an analysis tool within the framework of the acoustic monitoring of ALS speakers. More data are necessary for further investigations, however.

References

[1] F.L. Darley, A.E. Aroson, and J.R. Brown. *Motorspeech disorders.* Saunders, Philadelphia, PA, 1975.

[2] R.D. Kent, G. Weismer, J. Kent, and J.C. Rosen-

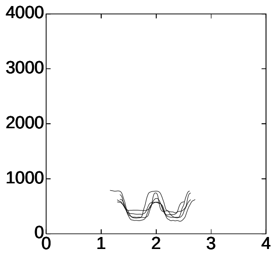
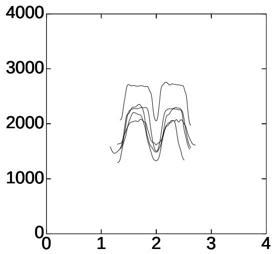
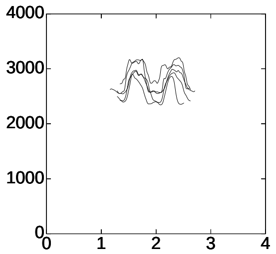
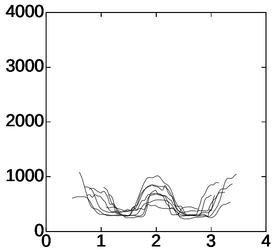
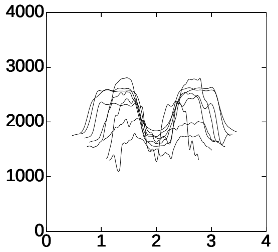
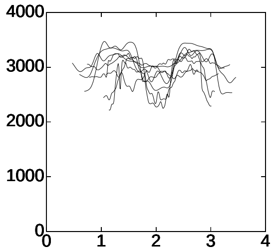
	F_1	F_2	F_3
Normals			
Dysarthrics			

Table 1: Changes over time of the first three formant frequencies of an [aiaia] segment sequence produced by normal and dysarthric speakers. The vertical axes represent formant frequencies (in Hz), and the horizontal ones time (in seconds).

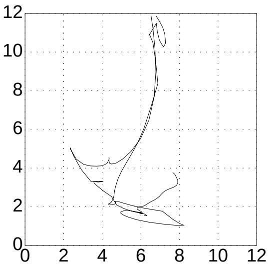
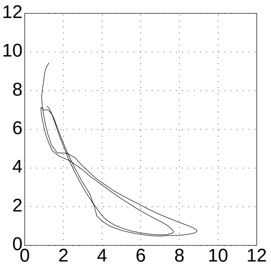
	Dysarthric	Normal
A_3&A_6		

Table 2: The trajectories of section pairs A_3 & A_6 of a ALS and a normal speaker inferred via formant-to-area mapping. The horizontal axes represent cross-section A_3 (in cm^2) and the vertical ones represent cross-section A_6 (in cm^2).

bek. Toward phonetic intelligibility testing in dysarthria. *J. Speech Hear. Dis.*, 54:482–499, 1989.

[3] D.H. Klatt and L.C Klatt. Analysis, synthesis and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Amer.*, 87(2):820–857, 1990.

[4] M. Mrayati, R. Carré, and B. Guérin. Distinctive regions and modes: A new theory of speech production. *Speech Comm.*, 7:257–286, 1988.

[5] J. Schoentgen and S. Ciocea. Direct calculation of the vocal tract area function from measured formant frequencies. In *Eurospeech*, volume 1, pages 745–748, 1995.

[6] J. Schoentgen and S. Ciocea. Kinematic acoustic-to-geometric mapping. In *ICPhS*, volume 2, pages 194–197, 1995.

[7] J. Schoentgen and S. Ciocea. Kinematic formant-to-area mapping. *Speech Comm.*, 4, 1997.

[8] J. Schoentgen and S. Ciocea. Post-synchronization via formant-to-area mapping of asynchronously recorded speech signals and area functions. In *Proceedings of the 5-th European Conference on Speech Communication and Technology*, Rhodes, Greece, 1997.

[9] B.H. Story, I.R. Titze, and E.A. Hoffman. Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Amer.*, 100:537–554, 1996.

[10] G. Weismer, K. Tjaden, and R. Kent. Can articulatory behanior in motor speech disorders be accounted for by theories of normal speech production? *Journal of Phonetics*, 23:149–164, 1995.

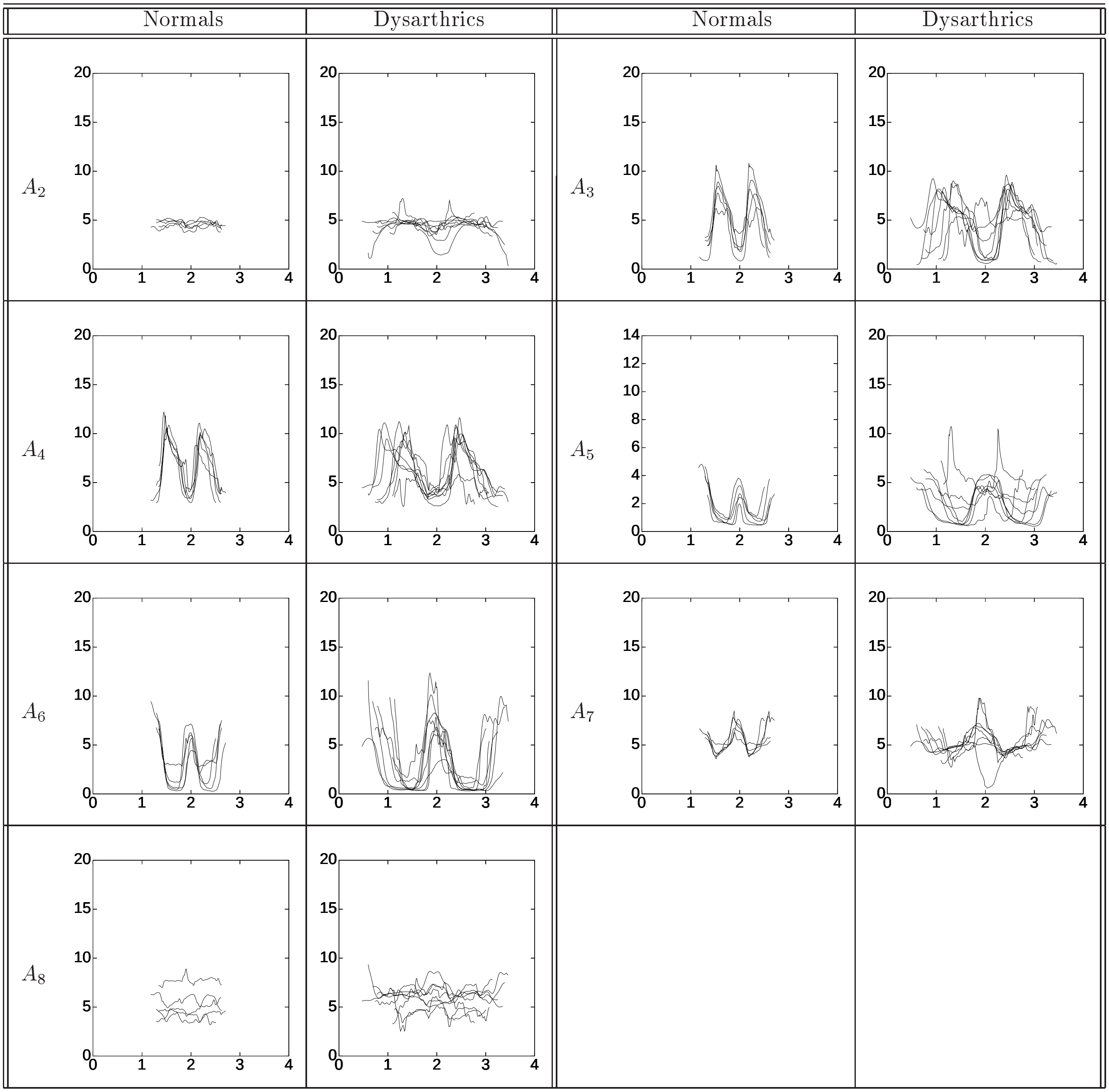


Table 3: Movements of cross-sections A_2 to A_8 (from top to bottom) obtained by means of formant-to-area mapping via the formant movements displayed in Table 1. The vertical axes represent cross-sections (in cm^2), and the horizontal ones time (in seconds).