TESTING VARIOUS METRICS FOR THE DESCRIPTION OF VOWEL DISTORTION IN DYSARTHRIA

Cécile Fougeron & Nicolas Audibert

Laboratoire de Phonétique et Phonologie, UMR7018 CNRS/Univ. Sorbonne Nouvelle, Paris, France
cecile.fougeron@univ-paris3.fr; nicolas.audibert@gmail.com

ABSTRACT

This study compares different acoustic metrics suitable for characterizing distortions in the articulation of vowels in the speech of dysarthric speakers with ASL. Metrics are tested for their performance in distinguishing patients vs. control speakers, and according to their relationship with speech intelligibility and perceived vowel alteration. Results support the need for different acoustic metrics in order to capture the large inter-speaker and inter-sex variation observed, and to reflect the various types of alteration possible.

Keywords: dysarthria, vowel, ASL, intelligibility, reduction

1. INTRODUCTION

Dysarthria refers to a group of neurologically-based motoric speech disorders, in which the impaired ability to control speech movements may result in alterations at any level of the speech production system. The assessment of dysarthria is often subjective and there is a need to develop reliable and quantitative measurements that could serve to monitor the presence, evolution, and severity of speech disorders, and/or treatment efficacy. Acoustic analysis of the patients' speech has great potential in addressing this issue (see [3]). However, when confronted with the multi-dimensionally impaired speech signal, one is also confronted with the issue of defining appropriate measurements.

The objective of this study is to evaluate several metrics suitable for the description and quantification of alterations in the articulation of vowels. We compare their performance in discriminating dysarthric patients from healthy controls, and we evaluate how they relate to perceived impairments in terms of intelligibility and vowel articulation accuracy.

The patients chosen for this study suffer from Amyotrophic Lateral Sclerosis (ALS). Their dysarthria is referred to as mixed (spastic and flaccid). It is associated with damage in both the peripheral and central motor systems. Speech impairments are caused by deficits in the magnitude, rate and strength of articulatory movement, which strongly affect the tongue and the orofacial musculature.

Impairments in the articulation of vowels in ASL patients have been reported both in perceptual and acoustic studies. In [2], the highest perceptual severity ratings for vowel articulation are found in the ASL group as compared to other dysarthria groups. Abnormal F1 and F2 formant frequencies, reduced formant transitions, and overall reduction of the vowel space area have been reported in ASL [6, 7, 8]. Reduction in vowel space area has also been found to correlate with perceived intelligibility of the patients [7, 8].

However, the acoustic metrics used thus far have not always been able to discriminate patients form healthy controls, or have failed to correlate with perceived speech intelligibility, e.g. [1, 6]. The comparison between patients and controls in the study of dysarthria is often challenged by the particularly large inter-speaker variation in dysarthric populations. If distortion in vowel articulation can be expected from the restricted mobility of the tongue, the severity and types of articulatory deficits, as well as possible compensation strategies, are speaker-dependent. Reductions can occur for instance in F1 and/or F2 dimensions, according to the vertical and/or horizontal restrictions in tongue movements and possible jaw compensation. The perceptual consequences of the vowel distortions on speech intelligibility are also not straightforward. In a reduced vowel space area, contrast between vowels can be preserved if there is little overlap.

In order to describe as accurately as possible distortions which arise in dysarthric vowel articulation, different acoustic dimensions have to be captured. The metrics tested in this study therefore relate to various possible reductions in articulatory/acoustic dimensions (F1, F2 and overall acoustic space), to centralization of articulatory targets toward a neutral vocal tract, and to neutralization of vocalic contrasts (overlap between the distribution of vowel categories).

2. METHOD

2.1. Population and speech material

The study included 27 French speakers diagnosed with ASL: 16 female (FP) and 11 male patients (MP). Patients were chosen to represent various levels of dysarthria severity, as judged by a group of 13 experts. The female group included 5 speakers with mild, 5 with moderate, and 6 with severe dysarthria. The male group included 2 speakers with mild, 4 with moderate, and 5 with severe dysarthria. The MP and FP groups were compared to age-matched healthy controls, with 16 female (FC) and 11 male controls (MC).

Patients and controls were recorded reading a story (about 200 words). A perceptual evaluation of the speech samples was done by 13 expert judges along various speech dimensions (rated on a 5-point scale with 4 = severely impaired). We will only report here on the score for the items 'global intelligibility' and 'vowel imprecision'.

2.2. Acoustic analysis & metrics

Ten to 12 occurrences of the vowels /i, e, a, o, u/ were selected in the text in an effort to control, as far as possible, for segmental context. A total of 2992 vowels were segmented manually, and F1 and F2 formants values were measured in their center and converted to bark for normalization.

The various metrics compared are presented in Table I. Metrics 1 to 5 characterize the distribution of the vowels in the acoustic space defined by F1 and F2. Metrics 6 and 7 refer to reduction in F1 or F2, which can be interpreted respectively as impairment in the range of high/low or front/back tongue movements (and rounding for the F2 of back vowels). Metrics 8 and 9 relate to overlap between vowels and thus to phonetic contrast. Some of these metrics are taken from the literature, while some are adapted to our French corpus. They also vary in the number of vowel categories required for their computation (from 2 to 5).

The values of all metrics were transformed into z-scores for correlation analyses.

Table I: Description of the 14 acoustic metrics tested (8 contains 6 measures).

1) Triangular vowel space area (tVSA)	Given F1 & F2 mean values of /i, a, u/; tVSA= $0.5*ABS(F1_i(F2_a-F2_u) + F1_a(F2_u-F2_i) + F1u(F2_i-F2_a))$ (see e.g. [8])
2) Pentagon vowel space area (pVSA)	Given F1 & F2 mean values of /i,e,a,o,u/, $pVSA=0.5*ABS(F1_i(F2_e-F2_u)+F1_u(F2_i-F2_e)+F1_e(F2_u-F2_i))+ABS(F1_u(F2_e-F2_o)+F1_o(F2_u-F2_e)+F1_e(F2_o-F2_u))+ABS(F1_a(F2_e-F2_o)+F1_o(F2_a-F2_e)+F1_e(F2_o-F2_a))$
3) Formant Centralization Ratio (FCR)	Ratio between formant values supposed to increase with centralization and values supposed to decrease with centralization; $FCR=(F2_u+F2_a+F1_i+F1_u)/(F2_i+F1_a))$ (see e.g. [4])
4) Custom FCR (cFCR)	Adaptation of FCR to the 5 vowels examined in this study and exclusion of $F2_a$ (not expected to vary with centralization in French); $cFCR=(F2_u+F1_i+F1_u+F2_o)/(F2_i+F1_a+F2_e)$
5) Dist_centroid	A common measure of centralization: Average distance of the 5 vowels centroid to the centroid of the speaker's vowel space
6) F2 range ratio (F2RR)	Designed to reflect the mobility of the tongue in the front-back dimension; $F2RR= F2_i/F2_u$ (see e.g. [4])
7) F1 range ratio (F1RR)	Analog of F2RR for description of alteration in tongue height dimension; $F1RR=F1_a/mean(F1_i, F1_u)$
8) V_1/V_2_overlap (measures for 6 vowel pairs)	For a given formant Fn and given that V_1 has a lower values than V_2, overlap between vowel pairs (V_1/V_2) close to each other in F1 or F2. Overlap is computed for i/e, e/a, a/o, u/o; i/u, e/o pairs. E.g., i/e_overlap (in F1 dimension)= $(F1_i+\sigma(F1_i))-(F1_e-\sigma(F1_e))$
9) totalVoverlap	Average of the 6 overlap measures between i/e, e/a, a/o, u/o; i/u, e/o pairs.

3. RESULTS

3.1. Predicting perceived alteration from acoustic metrics for the patient group

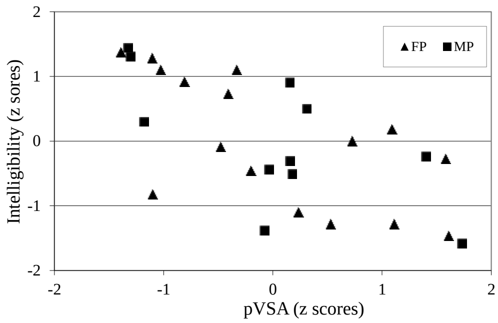
Perceptually, a strong relationship (|r|=.9 for both sex groups) is found between the scores on the intelligibility (Int) and vowel imprecision (VImp) items. Alterations in vowel production therefore contribute greatly to global perceived intelligibility.

Acoustically, among the metrics used to quantify vowel alteration, only a few predict the perceptual scores (either Int or Vimp), and variation is found between sex groups:

- For the male patients, 3 of the 14 acoustic metrics are significantly correlated with Int, showing a trend for less intelligible patients to have a smaller vowel space (pVSA*Int |r|=.7) and thus smaller distance from the vowels to the centroid of the space (Dist-centroid*Int |r|=.6), and more overlap in the F2 dimension between /e/ and /o/ (o/e_overlap*Int |r|=.6). The relationship between pVSA and intelligibility is illustrated in Figure 1. The pattern shown is similar to that observed between pVSA and VImp (|r|=.7): male patients with vowels perceived as more distorted show a smaller vowel space. Surprisingly, pVSA is the only acoustic metric correlated with VImp.

- For the female patients, less intelligible speakers show a smaller vowel space (pVSA*Int |r|=.7), with centralized vowels (cFCR*Int |r|=.5), and reduction in the opening dimension assessed by F1 (F1RR*Int |r|=.5). Reduction in the F1 dimension and of the vowel space area pVSA also predict the scores of VImp (|r|=.6 in both cases).

Figure 1: Distribution of pVSA according to perceived intelligibility for male and female patients.



3.2. Distinguishing patients from controls

One-way ANOVAs were performed for the male and female groups separately to test for an effect of speaker category (patient vs. control) on the acoustic metrics. For the female speakers, only one metric, e/o_overlap, was marginally significant (p=.05), with patients showing more overlap between the two mid-vowels in the F2 dimension than the control speakers.

For the male group, on the contrary, a significant speaker category effect was found for all metrics, except for u/o_overlap. Compared to control speakers, male patients have a significantly smaller vowel space area (t&pVSA), larger formant centralization ratios (FCR & cFCR), reduced F1 and F2 ratios (F1RR, F2RR), and more overlap between vowel pairs in F1 and F2 (except for the u/o pair), and in average over all pairs (totalVoverlap). Comparison of effect sizes was done by comparing η^2 values. Large effect sizes (η ≥40%) are found for F2RR, i/u_overlap, Dist_centroid, FCR, totalVoverlap, tVSA, and pVSA, and moderate to large effects are found for cFCR, F1RR, and 4 vowel pairs V_overlap (i/e, e/a, o/a, o/e).

4. DISCUSSION

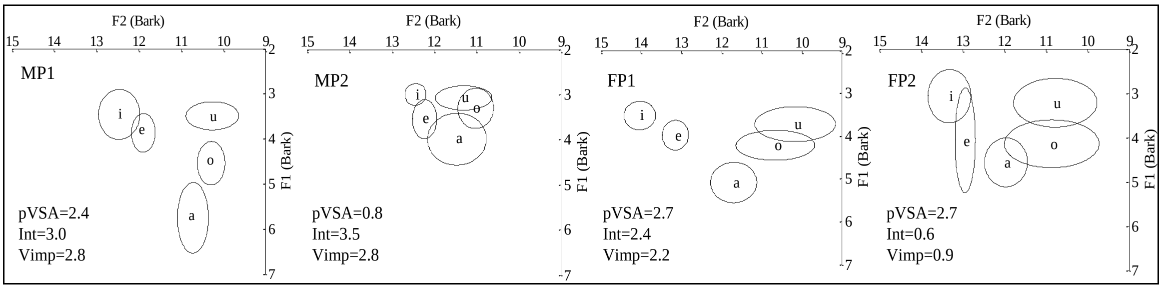
The aim of our study was to evaluate the suitability of various acoustical metrics for the description of vowel alterations in patients with ASL. Conclusions about the performance of the metrics appear to be purpose- and population-dependent. Indeed, results differ according to whether an acoustic metric is tested for its performance in discriminating impaired vs. normal speech, or for its relationship with the perceptual evaluation of the impairments.

In section 3.2, when an effect of speaker category is found (i.e. for the male population), most of the metrics robustly differentiate dysarthric patients from healthy controls. Comparison of the effect sizes indicates that differences in vowel articulation between patients and control speakers are best accounted for by the 2 vowel space area metrics (pVSA & tVSA), the 2 centralization measures (FCR & Dist_centroid), 2 metrics related to reduction in the front-back tongue dimension (F2RR, i/u_overlap), and the global measure of overlap between vowel pairs (totalVoverlap). In section 3.1, however, not all of these metrics are found to correlate with the perceived speech distortions. For example, reduction and centralization in the vowel acoustic space are also found to predict perceived intelligibility, but only when the respective metrics are computed with 5 vowels (pVSA, cFCR, Dist-centroid) and not 3 vowels (tVSA, FCR). This suggests that it is worth including more than 3 vowel categories in the computation of acoustic metrics.

Overall, the results concerning the relationships between acoustics and perception were unexpected. Since the item ‘vowel imprecision’ was specifically selected to target the perceptual judgments toward these segments, we expected more and stronger correlations with acoustic measures. Only pVSA was significant for both sex groups, and F1RR for the female patients. It is possible that the vowels selected for acoustic analysis were not good exemplars of the impaired speech signals. It is also probable that our metrics based on F1 and F2 dimensions are not able to capture the whole range of possible vowel alterations, and particularly those linked to hypernasality or impaired voice quality that are frequent in ASL.

Even if vowel alteration is found to be a good predictor of the perceived global intelligibility of the speech files, it is obviously not its only predictor. Three to 4 acoustic metrics are found to predict intelligibility scores according to sex groups. As previously mentioned in the literature, reduction in intelligibility goes with reduction of the vowel space (pVSA), and a global shift of vowel targets toward the space center (Dist_centroid, cFCR). It is also a function of alterations in specific dimensions of the space (e.g. F1RR, e/o_overlap). Nonetheless, the relationship between perceived alteration and acoustic realization is not as straightforward as assumed by these global tendencies. Figure 2 illustrates the acoustic realization of the vowels for four selected patients, and gives their pVSA and Int and Vimp scores. While speaker FP1 and FP2 display the same overall reduced space area (pVSA), they are rated very differently in perception (for both Int and Vimp). The reverse is found for MP1 and MP2: while they have similar perceptual scores, they have drastically different vowel spaces.

Figure 2: Vowel spaces with acoustic (pVSA) and perceptual (Int and VImp) descriptors for 2 male (MP) and 2 female patients (FP).



Interestingly, our results show that sex- or speaker-dependent patterns of reduction have to be taken into account, and the acoustic metrics to be used need to capture the full range of variation. The examples given in Figure 2 are good illustrations of the various patterns that can be observed. Speaker MP2 presents a strongly reduced and centralized acoustic space with relatively little overlap in the realization of /i, e, a/. On the contrary, speaker FP2 shows a larger vowel space but with much variation in the realization of each vowel and therefore more overlap for these 3 vowels. Speakers FP1 and MP1 also show different patterns of reduction that may be interpreted in terms of restriction in tongue mobility: the space is compressed in the F1 dimension for FP1 and probably reflects restriction in tongue height movements, while the space is compressed in the F2 dimension for MP1, suggesting restriction in front-back movements.

Finally, the sex-group difference observed in 3.2 for the patients vs. controls discrimination also raises questions about speaker selection. In ASL, heterogeneity in the patient profiles is well known and is linked to variation in neural degeneration, in disease progression and severity, and sex [5]. The speakers with severe dysarthria in our female group were not judged as severe as the ones in the male group. Therefore, the overall severity of the female group may not have been strong enough to differentiate them as a group from the healthy controls. Moreover, a qualitative examination of the vowels of the female control group revealed that more female control speakers had reduced/centralized/overlapping vowels than did the male controls. Further investigation is needed to understand this aspect in a study with a larger population, a larger range of severities, and other neurological deficits.

To conclude, our findings and the variation we have observed when looking at the patients individually suggest that a detailed description of the impairment of vowel articulation in dysarthric patients has much to gain from the use of various types of acoustic metrics covering all dimensions subject to alteration.

5. ACKNOWLEDGEMENTS

This study is funded by the ANR-08-BLAN-0125.

6. REFERENCES

[1] Bunton, K., Weismer, G. 2001. The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *JSLHR* 44, 1215-1228.
[2] Darley, F., Aronson, A., Brown, J. 1969b. Clusters of deviant speech dimensions in the dysarthrias. *JSHR* 12, 462-496.
[3] Kent, R., Weismer, G., Kent, J., Vorperian, H., Duffy, J. 1999. Acoustic studies of dysarthric speech: Methods, progress, and potential. *J. Comm. Disorders* 32, 141-186.
[4] Sapir, S., Ramig, L., Spielman, J., Fox, J. 2010. Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. *JSLHR* 53(1), 114.
[5] Strand, E., Buder, E., Yorkston, K., Ramig, L. 1994. Differential phonatory characteristics of four women with amyotrophic lateral sclerosis. *J. Voice* 8, 327-339.
[6] Turner, G., Tjaden, K. 2000. Acoustic differences between content and function words in amyotrophic lateral sclerosis. *JSLHR* 43(3), 796-815.
[7] Turner, G., Tjaden, K., Weismer, G. 1995. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *JSHR*. 38, 1001-1013.
[8] Weismer, G., Jeng, J.Y., Laures, J.S., Kent, R.D., Kent, J.F. 2001. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica & Log.* 53(1), 1-18.