ACOUSTIC METRICS OF VOWEL ARTICULATION IN PARKINSON'S DISEASE: VOWEL SPACE AREA (VSA) Vs. VOWEL ARTICULATION INDEX (VAI)

S. Sapir\(^1\), L. Ramig\(^2,3\), J. Spielman\(^2,3\), C. Fox\(^2,3\)

\(^1\)Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel
\(^2\)Department of Speech, Language and Hearing Sciences University of Colorado, Boulder
\(^3\)National Center for Voice and Speech-Denver

Abstract: Acoustic analysis of speech is a powerful, noninvasive, and cost effective tool to study different aspects of motor speech disorders such as the dysarthria associated with PD. In this presentation, we will discuss the rationale for using acoustic analysis, its advantages and disadvantages, and methods to overcome these disadvantages. As an example, we will address the use of vowel space area (VSA) in the study of dysarthric vowel articulation in PD. Although the VSA is theoretically driven, it is highly sensitive to inter-speaker variability, which, statistically speaking, introduces noise. This noise can mask important differences that do exist between speakers with and without PD. Some of this noise can be reduced by logarithmic transformation of the formant frequencies. However, even with this transformation, some statistical noise might be still present. Recently, Sapir and colleagues introduced two acoustic metrics -- the Vowel Articulation Index (VAI) and its inverse, the Formant Centralization Ratio (FCR) -- that are theoretically driven and empirically tested. These metrics show promise as they effectively reduce inter-speaker variability noise while maintaining high sensitivity to vowel centralization (the latter reflecting abnormally reduced (hypokinetic) articulatory movements in PD). Data will be presented of 38 individuals with Parkinson’s disease and 14 healthy controls whose speech was effectively differentiated by the VAI, but not the VSA, yet the logarithmically scaled VSA (LnVSA) did significantly differentiate between dysarthric and normal speech, although not as strongly as the VAI.

Keywords: Parkinson disease, acoustic analysis, speech

I. INTRODUCTION

Individuals with Parkinson's disease (PD) often suffer from hypokinetic dysarthria, a neuromuscular disorder of voice and speech, resulting and characterized by reduced vocal loudness, monotone voice, and imprecise consonants and vowels. Most types of dysarthria, including that associated with PD, are characterized by articulatory undershoot, i.e., reduced range of articulatory movements, to the extent that the intended place and degree of vocal tract constriction are not fully achieved. This undershoot is likely to result in vowel formant centralization; i.e., formants that normally have high center frequencies tend to have lower frequencies, and formants that normally have low center frequencies tend to have higher frequencies [1,2].

A common way to represent this centralization is with the VSA [3]. In English, the VSA is usually constructed by the Euclidean distances between the F1 and F2 (frequency) coordinates of the corner vowels /i/, /u/, and /a/ (triangular VSA), or the corner vowels /i/, /u/, /a/, and /ae/ (quadratic VSA) in the F1-F2 plane. The formula of the VSA constructed with the vowels /i/, /u/ and /a/, is $\text{ABS}((F1_i \times (F2_u - F2_i) + F1_a \times (F2_u - F2_a) + F1_u \times (F2_i - F2_a))/2)$.

Due to articulatory undershoot and subsequent centralization of vowels, the VSA in the speech of individuals with dysarthria is expected to be compressed relative to that of normal speech (Kent & Kim, 2003). Improvement in speech due to natural recovery or treatment effects should be reflected in the expansion of the VSA toward normalcy (e.g., Sapir et al, 2003). Although several studies demonstrated the ability of the VSA to differentiate between dysarthric and normal speech and to monitor treatment effects (e.g., Liu et al., Sapir et al, 2003), other studies failed to do so, even though a trend toward centralization of vowels was evident (e.g., Weismer et al., 2001). The reasons for the inconsistent performance of the VSA are not clear, although interspeaker variability appears to be a major factor. Interspeaker variability in vowel formant frequencies and VSA is expected due to numerous factors (cf. Sapir et al., 2010), the most obvious of which are anatomical differences, such as those associated with gender and age (e.g., size and shape of the vocal tract).

It is clear that to improve differentiation between dysarthric and normal speech, the acoustic metric must be minimally sensitive to interspeaker variability and maximally sensitive to vowel formant centralization.
Recently, Sapir introduced two acoustic metrics that are designed to be minimally sensitive to interspeaker variability and maximally sensitive to vowel formant centralization. These metrics include the Vowel Articulation Index (VAI), expressed as \((F2i+F1a)/(F2u+F2a+F1u+F1i)\), and its inverse, the Formant Centralization Ratio (FCR), expressed as \((F2u+F2a+F1u+F1i)/(F2i+F1a)\). (Sapir et al., 2006; Sapir et al., 2010). Note that in the VAI the numerator is likely to decrease and the denominator is likely to increase with vowel formant centralization, whereas in the FCR the numerator is likely to increase and the denominator to decrease with vowel centralization. Importantly, at least in American English, the normal VAI values should be close to 1.0, as the sum of formant frequencies in the denominator is very similar to the sum of formant frequencies in the numerator. Thus, the VAI may be considered a function that normalizes the relationships between the vowels across speakers. The purpose of the present study is to demonstrate the ability of the VAI, VSA, and LnVSA, to differentiate between normal and abnormal vowel articulation. We predicted that the VAI will perform best and the VSA worst. We also predicted that the LnVSA will perform better than the VSA because logarithmic scaling of formant frequencies tend to reduce interspeaker variability.

II. METHODS

Subjects. The subjects in this study participated in our previous study (Sapir et al., 2010). They all spoke American English as their first language and the majority of them resided in Tucson Arizona or Denver Colorado. Of these individuals, 38 had idiopathic Parkinson's disease (PD) (19 M, 19 F) with dysarthria of different levels of severity, and 14 individuals (7 M, 7 F) served as healthy, age-matched and gender-matched controls (HC). The VAI and VSA were constructed from the frequencies of the first (F1) and second (F2) formants of the vowels /i/, /u/, and /a/. These frequencies were also logarithmically scaled for the construction of a logarithmic version of the VSA (henceforth, LnVSA). The vowels /i/, /u/, and /a/ were extracted from the phrases “The blue spot is on the key,” “The potato stew is in the pot” and “Buy Bobby a puppy” (target words: “key”, “stew”, and "Bobby") or the phrase “The stew pot is packed with peas” (target words "stew", "pot", "peas"), with several repetitions of each of the phrases. Details of the recordings and acoustic analysis are described elsewhere (Sapir et al., 2010).

III. RESULTS

The main findings of this study are summarized in Table 1. The table shows the means and SDs for the two groups (PD, HC) and for three acoustic metrics (VSA, LnVSA, VAI), as well as t-test results and p values for significance. Also, the coefficient of variation (CV) for the two groups and three metrics (VSA, LnVSA, VAI) are shown at the bottom of the table. Effect size (ES) measures (Cohen, 1988) are also used to indicate the clinical significance of the differences between the two groups the degree. In general, a value of 0.80 and higher indicates highly significant differences between the two groups. A value of 0.50 a medium effect and a value of 0.20 indicates a small or a negligible effect. As can be seen, the VSA does not significantly differentiate between the two groups (PD vs. HC). The LnVSA improves performance considerably, whereas the VAI performs best, both statistically and in terms of a large effect size.

### Table 1. The ability to differentiate between the dysarthric and normal vowel articulation by the VSA, LnVSA, and VAI. CV = Coefficient of Variation ; ES = Effect Size (>0.8 large, 0.5 medium, 0.2 small effect).

<table>
<thead>
<tr>
<th></th>
<th>VSA (Hz)</th>
<th>LnVSA (LnHz)</th>
<th>VAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD ([n=38]) Ave=</td>
<td>232120</td>
<td>0.23</td>
<td>0.96</td>
</tr>
<tr>
<td>SD=</td>
<td>(96155)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>HC ([n=14]) Ave=</td>
<td>279524</td>
<td>0.28</td>
<td>1.05</td>
</tr>
<tr>
<td>SD=</td>
<td>(68810)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>t-test</td>
<td>p=</td>
<td>0.0579</td>
<td>0.0099</td>
</tr>
<tr>
<td>ES=</td>
<td>0.57</td>
<td>0.77</td>
<td>1.24</td>
</tr>
<tr>
<td>PD</td>
<td>CV=</td>
<td>41%</td>
<td>33%</td>
</tr>
<tr>
<td>HC</td>
<td>CV=</td>
<td>25%</td>
<td>24%</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

These findings strongly suggest that by reducing interspeaker variability and by maximizing sensitivity of the acoustic metric to the differences between normal and abnormal speech one can improve the reliability and validity of the acoustic analysis. Our task is to do the same for other acoustic metrics of speech.

Once we have acoustic metrics that comply with these two criteria, we can combine the different acoustic
metrics and use more sophisticated analyses to differentiate between normal and abnormal speech and to monitor changes associated with disease progression and treatment effects. Finally, we addressed only one issue related to improving speech signal processing for clinical and research processes. There are other important factors that we should consider, such as the problem of recording speech in a noisy environment, using inappropriate recording equipment and procedures, and choosing the wrong speech tasks to elucidate and register the speech abnormalities in PD.

V. CONCLUSIONS

Unlike the VSA, the VAI is a powerful acoustic metric to reduce interspeaker variability and enhance sensitivity to dysarthric vowel articulation.

REFERENCES


