Pitch analysis of dysarthria helps differentiating between dysarthria mechanisms

Horia-Nicolai Teodorescu

Horia-Nicolai Teodorescu - Institute of Computer Science of the Romanian Academy –Iasi Branch, "Gheorghe Asachi" Technical University, 67

Dedicated to the memory of the late Professor Petre Brânzei (1916–1985)

ABSTRACT

This brief preliminary article suggests the use of pitch analysis for gaining insight in the mechanisms of dysarthria affecting specific subjects, when dysarthria involves poor control of the vocal cords. Examples are provided.

KEYWORDS:

Speech technology, dysarthria, dysarthria mechanism, vocal cords, pitch analysis.

INTRODUCTION

Recent progresses in medicine, on one side, and technology, on the other side, approached these seemingly parallel domains to the point of joining them. Almost every research field has recently benefitted of information technology in multiple ways. Medicine currently uses extended signal and information processing in ways unconceivable 30 years ago, from apparently simple personal health monitors to diagnostic expert systems and robotic surgery. It may seem that the realms of psychiatry and psychology lag behind, but recent

applications of information processing, often pushed by lucrative uses such as marketing and hiring analytics and homeland security have pushed the use of computer science into these domains. During the last decades, the automatic recognition of emotions and sentiments in speech and writing became an established domain in speech and natural language processing technologies. Also, software tools revolutionized pre-clinical analysis and medical diagnostic, including those in neurologic and psychiatric disorders. Voice technology includes methods and means (computer programs packages) for the

analysis of voice signals, allowing the detection and graphical representation of key parameters of the voice such as pitch, formants, frequency spectrum, and amplitude variation, among many others.

The pitch, also named ‘zero formant’, is the basic ‘tone’ (frequency component) in voice signal. It is produced by the vibrations of the vocal cords and thus provides essential information on the vocal cords operation and on their neurological control. The pitch value depends on the age (higher for children), sex (typically higher for women), emotion, and prosody. Its value variation is one of the major elements characterizing prosody (1, 2, 3, 4).

Fourier analysis, also named frequency or spectrum analysis, reveals the formant structure of the voiced sounds. The formants are the major maxima in the spectrum and correspond to resonances of the vocal tract. The formant values are modified by controlling the shape of the elements of the vocal tract during speech. Because the parts of the vocal tract are controlled in different areas of the brain than the vocal cords, the changes in the formants, the pitch control, and the duration of the sounds (e.g., Hartelius (5)) offer different information on the brain operation (6, 7, 8, 9). In addition, pitch, the durations of the phonemes and the formants are influenced by the mood, health state, and the emotions. Depression translates into specific emotional states, including sadness, and into variations of the state mental from the “normal”. In turn, these changes translate into involuntary variations of speech features (characteristics). These variations of speech features can be automatically detected by appropriate signal processing means, information extraction, specific classification techniques, and automatic decision-making processes (1, 2, 3, 4).

The main contribution of this study is to draw attention on the usefulness of the variability of the pitch, expressed as the standard deviation, denoted by σ , of the pitch divided by its average, $\overline{F0}$, that is $\sigma/\overline{F0}$, in assessing the speech impairment in dysarthria and in apraxia of speech (AOS).

DYSARTHRIA AND DYSARTHRIC SPEECH

Speech disturbances may be the result of lack of muscular coordination (ataxia) and may appear with various intensities (from mild to severe) in traumatic brain injury, progressive supranuclear palsy, Parkinson’s disease, Wilson’s disease, cerebral palsy, multiple sclerosis, Huntington’s disease, hydrocephalus, stroke, brain tumor, Freidreich’s ataxia, cerebellar ataxia, and spinocerebellar ataxia(10). ‘Apraxia’ (with reference to speech disorders), or ‘apraxia of speech’ (AOS) is a term introduced in the 1960s for naming a set of motor speech disorders due to a large number of causes, including but not limited to aphasia or dysarthria (11). Ogar et al. (11) recall that apraxia of speech (AOS) “*is often the first symptom of neurodegenerative diseases, such as primary progressive aphasia and corticobasal degeneration*” and reviewed the “*articulation and prosody deficits associated with AOS*”.

Most papers on the topic of dysarthria focused on speech task intelligibility and speech rate in dysarthria. Hartelius et al.(5) analyzed the temporal characteristics of dysarthric speech in multiple sclerosis (MS) and of general ataxic dysarthria and found that in ataxic dysarthria and MS the syllables have “*significantly increased durations and decreased intrautterance variability ... as well as significantly increased interutterance variability*.” Weismer et al. (12) analyzed the speech for speakers with amyotrophic lateral

sclerosis (ALS) and Parkinson's disease (PD). They found that *“temporal variables typically differentiated the ALS group, but not the PD group, from the controls, and that vowel spaces were smaller for both neurogenic groups as compared to controls...”* Lansford and Liss (13) determined that typical vowel metrics are good predictors of dysarthria; they emphasize that *“only the vowel metrics that captured slope of the second formant (F2) demonstrated between-group differences across the dysarthrias.”*

Vowel pronunciation was the focus of the studies of several groups. Fougerson and Audibert (14) investigated the discrimination power of various metrics for the characterization of vowel distortion in dysarthria. Several studies, for example by Skodda, Visser, and Schlegel (15) and by Whitfield and Goberman (16) focused on vowel articulation and vowel space in Parkinson's disease. The so-called vowel space is simply the area of the polygon determined in the plane of the first two formants (F1 and F2) by the ‘average’ vowels pronounced by a subject. The analysis of formant centralization was found effective in explaining the voice quality ratings by listeners (perceptual rating) by Fletcher et al. (10). Basically, the vowel centralization is computed as a function of the values of the first two formants for a specified subgroup of vowels. For example, (13) use the metric named VSA defined as:

“Vowel centralization ... VSA ... is calculated using F1 and F2 of the (v:), (i:), and (o:) vowels... averaged for each speaker ... by plotting these values as coordinates in an F1/F2 plane and calculating the resulting triangular area.”

For speech analysis, Fletcher et al. (10) use Praat (17) software package, also used in this study.

APPLICATION OF VOICE ANALYSIS TO DYSARTHRIA

Material and method

In this brief study we use WASP™ program (WASP speech analyzer) (18), authored by Mark Huckvale, University College London and UCL Phonetics and Linguistics, Faculty of Brain Sciences. In addition, we use Praat program (software package) described in (17). We continue the line described in (19) of making use of and analyzing resources and cases available on the Internet. We use for exemplification the three dysarthria cases in a video made accessible by Mayo Clinic (20). The waveform, sonogram and pitch trace of the entire sound file are shown in Fig. 1.

Analysis results

In minor cases of dysarthria, such as childhood apraxia of speech, the impairment consists only in producing a few phonologic errors (wrong sounds) (20). The first case discussed falls in this category.

Fig. 1. The entire recording of speech at (20). Upper panel: waveform. Second panel: sonogram, with coarse spectrum resolution. Third panel: sonogram, fine grain resolution. Lower panel: pitch trajectory. Graph obtained in WASP

Fig. 2. Speech of the first child in the recording, with minor inaccurate movements. Upper panel: waveform. Middle panel: sonogram, fine grain resolution. Lower panel: pitch trajectory. Graph obtained in WASP

For the child #1 in the recording (20), the trajectory of the pitch is quite smooth locally, see Fig. 2 and 3, yet it has large variations as for a normal prosody. Definitely, the graph of

the pitch for child #1 shows only minor speech impairment for her age (difficulties related to brief tremors in the pitch trajectory).

Fig. 3. Speech of the first child in Praat (segment slightly larger than in Fig. 2). Graph obtained with Praat

The second case analyzed (Child #2) is a dysarthria case with strong “distortion

sounds, spasticity, weakness, slow rate of speech, but the child has a good level of

intelligence” (20). The dysarthria is due to neurological (and potentially psychologic issues, which cannot be neglected in such situations), translated into inaccurate control

of the phonation. The waveform, sonogram and pitch trace are shown in Fig. 4 (in WASP) and in Figs. 5 and 6 in Praat.

Fig. 4. Speech of the second child. Upper panel: waveform. Middle panel: sonogram, fine grain resolution. Lower panel: pitch trajectory. Graph obtained in WASP

Fig. 5. Pitch trajectory, second child. Graph obtained with Praat.

Obviously, the neural control of the vocal cords is greatly deficient, because the prosody is almost absent, while the trace of the pitch is very noisy. Also notice in Fig. 6 the imprecise traces of the formants. It is interesting to contrast the pitch variation of this child with the pitch of the others and that of the

presenting doctor (her voice also appears in the same recording.) In Fig. 7, the speech segment of the second child is between the two vertical lines. It is clear that the pitch is noisy, yet its average is flat, with no prosody whatsoever.

Fig. 6. Second child. Waveform and formants. Graph obtained in Praat

Fig. 7. Second child. Upper panel: waveform. Middle panel: sonogram, fine grain resolution. Lower panel: pitch trajectory. Graph obtained in WASP.

There are several types of dysarthria; it is believed that they differ by the root-cause, the part of the brain involved in speech and operating unsatisfactorily. As a consequence, the type and characteristics of the respective speech produced in the various dysarthria types are very different too. The next case of dysarthria analyzed has hyper-nasality, that is, the air flux to the nose is exaggerated (20).

According to the presentation of the case in the Maya Clinic in the cited video, the speech rate is excessively slow. Notice that the original video on this case includes segments of pathologist voice; we carefully selected only parts corresponding to the child (male). Two segments of speech for this child are shown in Fig. 8.

Fig. 8. Voice of the child #3, dysarthria. Pitch trajectory. Graph obtained in WASP. Panels as for Fig. 4.

The pitch trajectory in Fig. 8 is both noisy and having jumps in value, due to the lack of control of the voice. In Fig. 8, the structure of the formants is difficult to find, meaning a lack of control of the vocal tract. On the other hand, the vocal fold control is, in some respect, better than for the previous child, with the pitch variation smoother, except the large jumps, when the control is lost.

For further clarifying the visible differences in the evolutions (trajectories) of the pitch for the three cases, we computed the average and the standard deviation of the values of the pitch for the three subjects, for the speech segments shown in the above figures. The results are given in Table 1. The first child, who has only a minor childhood apraxia of speech, is a good reference for comparison and for assessing the other two cases.

Table 1. Basic statistic features of the pitch for the three analyzed cases of dysarthria

	Child #1	Child #2	Child #3
Average pitch, $\overline{F0}$	171.5	148.4	95.8
STDEV pitch, σ	64.5	37.7	56.5
STDEV/average, $\sigma/\overline{F0}$	0.38	0.25	0.59
Comment	Almost normal	Almost no prosody, noisy variation	Large jumps in the pitch

Although the pitch of the second child is noisy, it has however little variability around the average value; therefore, its standard deviation and ratio of the standard deviation to average are the smallest. The pitch for the

third child has very large jumps; as a consequence, its standard deviation and ratio of the standard deviation to average are the largest.

DISCUSSION AND CONCLUSIONS

This study highlighted the usefulness of information processing tools for speech analysis in neurology and psychiatry. Previous studies of dyslexia favored the analysis of the formants, especially the study of the variation of the area of triangles in the plane of the first two formants for specified vowels. We suggest that the visualization and analysis of the pitch variation is essential for good differentiations between various types of dysarthria and for revealing the causes and main mechanism of dysarthria in specific subjects. We are not aware of such an approach in the literature. We also argue that the use of voice analysis should become a regular, information processing based tool in assessing the condition of the subjects in neurology and psychiatry.

ACKNOWLEDGMENTS AND DISCLOSURES

I thank Prof. PhD Philol. Călina Gogălniceanu for the encouragement of publishing in this journal. The author states no conflict of interests. The author retains the right to publish an extended version of this preliminary brief article.

REFERENCES

1. S. Howard, B. Heselwood. 2012, The Contribution of Phonetics to the Study of Vowel Development and Disorders. In Handbook of Vowels and Vowel Disorders, M.J. Ball, J.S. Damico, M.J. Ball, F.E. Gibbon (Eds.), Routledge, Accessed on: 20 Feb 2019, <https://www.routledgehandbooks.com/doi/10.4324/9780203103890.ch3>.
2. D. Jurafsky and J.H. Martin, Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd Ed. 2008. Prentice-Hall, Inc., New Jersey.
3. R.D. Kent, H.K. Vorperian, Static measurements of vowel formant frequencies and bandwidths: A review. J. Communication Disorders. 2018; 7474–97.
4. Teodorescu H.N. Fuzzy Logic in Speech Technology - Introductory and Overviewing Glimpses. In: Fifty Years of Fuzzy Logic and Its Applications. Book Series: Studies in Fuzziness and Soft Computing, Springer Int. Publ., 2015; 326:581-608.
5. Hartelius L., Runmarker B., Andersen O., Nord L., Temporal Speech Characteristics of Individuals with Multiple Sclerosis and Ataxic Dysarthria: 'Scanning Speech' Revisited. Folia Phoniatr Logop 2000; 52:228-238.
6. Teodorescu H.N., Feraru SM, A study on Speech with Manifest Emotions. Proc. 10th International Conference on Text, Speech and Dialogue, Pilsen, Czech R. Lecture Notes in Artificial Intelligence 2007; 4629:254-261.
7. Teodorescu H.N., Feraru S.M. Analyzing emotions in spoken Romanian. Proceedings of the Romanian Academy. 2007; 8(2):161-168.

8. Feraru S.M., Teodorescu H.N., Zbancioc M.D. SRoL - Web-based resources for languages and language technology E-learning. International Journal of Computers Communications & Control.2010;5(3):301-313.
9. Teodorescu H.N., Zbancioc M., Feraru S.M. Analysis of vowel triangle variation for the Romanian language related to emotional states. Proc. 10th International Symposium on Signals, Circuits and Systems (ISSCS).2011. DOI: 10.1109/ISSCS.2011.5978727.
10. Fletcher A.R., McAuliffe M.J., Lansford K.L., Liss J.M. Assessing Vowel Centralization in Dysarthria: A Comparison of Methods. Journal of Speech, Language, and Hearing Research.2017;60(2):341-354.
11. Ogar J., Slama H., Dronkers N., Amici S., Gorno-Tempini M.L. Apraxia of Speech: An overview, Neurocase. 2005;11(6):427-432.
12. Weismer G., Jeng JY., Lares J.S., Kent R.D., Kent J.F. Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. Folia Phoniatr Logop. 2001;53:1-18.
13. Lansford K.L., Liss J.M. Vowel Acoustics in Dysarthria: Speech Disorder Diagnosis and Classification, Journal of Speech, Language, and Hearing Research.2014;57(1):57-67.
14. Fougerson C., Audibert N. Testing Various Metrics for the Description of Vowel Distortion in Dysarthria, ICPhS XVII, Hong Kong, 2011.
15. Skodda S., Visser W., Schlegel U. Vowel Articulation in Parkinson's Disease. Journal of Voice. 2011;25(4):467-472.
16. Whitfield J.A., Goberman A. M. Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. J. Communication Disorders.2014;51:19-28.
17. Boersma P., Weenink D. Praat for Windows 64-bit edition: praat6046_win64.zip from www.fon.hum.uva.nl/praat/download_win.html, (Univ. of Amsterdam).2019.
18. Huckvale M. WASP™ program (wasp speech analyzer).University College London and UCL Phonetics and Linguistics, Faculty of Brain Sciences.
19. Teodorescu H.N., Gogalniceanu D. Voice-related symptom and knowledge-bases using internet mining. Proc. International IEEE Conference on Speech Technology and Human-Computer Dialogue (SpED), Bucharest, Romania.2017.
20. Mayo Clinic. Differentiating Childhood Apraxia of Speech (CAS) from other types of speech sound disorders. Video at <https://www.youtube.com/watch?v=WhuEM9tE-k>.

Correspondence:

Horia-Nicolai Teodorescu,
Institute of Computer Science of the Romanian Academy –Iasi Branch, "Gheorghe Asachi"
Technical University, 67, Bd. D. Mangeron, 70050 Iasi, Romania

Submission: 27 oct 2019

Acceptance: 25 jan 2020