Dysarthric Speech Enhancement using Formant Trajectory Refinement

Divya Das
Department of Computer Science and Engineering
Govt: College of Engineering Sreekrishnapuram, Palakkad, Kerala, India

Dr. C. Santhosh Kumar
Department of Electronics and Communication Engineering
Amrita School of Engineering, Amrita Nagar, Coimbatore, Tamil Nadu, India

Dr. P. C. Reghu Raj
Department of Computer Science and Engineering
Govt: College of Engineering Sreekrishnapuram , Palakkad, Kerala, India

Abstract- Dysarthria is a motor-neuro disorder that affects the quality of articulation required to produce speech. Also, there are temporal inconsistencies in the speech produced by people with dysarthria, leading to inconsistent formant trajectories. The trajectories change slowly in dysarthric speech, when compared to normal speech.
In this work, we refine the formant trajectories of the dysarthric speech to improve its intelligibility. We use the P.563, P.862, standards along with composite measures to evaluate the quality of speech, before and after the refinement; we used NEMOURS database for the experiments involving mild dysarthria.
For the proposed work, we attempt to emphasize the fast variations in the formant trajectories to enhance speech quality. It was observed that the quality of speech improved significantly. Our method will therefore encourage the dysarthric people to communicate more effectively and improve the pace of their rehabilitation. To the best of our knowledge, this type of work is not reported elsewhere.

Keywords – Formant frequencies, Intelligibility, Enhancement

I. INTRODUCTION

The term dysarthria is used to describe changes in speech production characterized by impairment in one or more of the systems involved in speech. The three major systems involved in speech production are respiration, voice production and articulation. Voice is produced by the larynx and the oral structures articulate to modify the sound source produced by the larynx. For instance, speakers may show a slow rate of speech, but particular words or phrases within that utterance may be produced with a rapid rate. The oral structures such as the tongue and lips are rigid, resulting in a reduced range of movement. This effectively dampens the speech signal and distorts the accuracy of the sound (consonant or vowel) production [8].
Clinically it can be cured by taking proper medicines according to the classes of dysarthria and by giving speech therapy to the people with dysarthria. In computational approach we can support them by improving the intelligibility of the dysarthric speech[1]. The intelligibility can be improved at the signal level by considering the speech of the dysarthric people[7]. The main challenges are their inability to control the pitch, timing, tone, strength, breathe while speaking. Acoustic characteristics of each phone are different and they have different spectral characteristics[13]. For people affected with dysarthria in the continuous speech these spectral changes do not happen as they should be with normal people[2]. As a result, their formant trajectories change very slowly. To make them closer to the speech of normal people this formant trajectory variations to made faster[3]. In this work we emphasize on a system that induces variations on the formant trajectories and re synthesize the new speech.

II. FORMANT TRAJECTORY REFINEMENT

A. Database Used

In this work we use nemours database for the dysarthric speech enhancement study. Nemours database contains the sentences of the pattern “The X is Y ing the Z” spoken by the dysarthric people in English. The X,Y and Z are the words with consonant- vowel-consonant(CVC) combination with maximum articulation problems in dysarthric people. Dysarthric patients, while using CVC combination patterns will take a long time to switch from the vowel to consonants.

Dysarthric speech enhancement using formant trajectory refinement study uses eleven directories with specific dysarthric data. Among them four directories has mild dysarthric speech, and this is the focus of this work. Each directory has 74 different sample consonant- vowel-consonant (CVC) combination data. Out of these, 37 original combination of X,Y and Z and another 37 sentences were generated by swapping the X and Z tokens in the original set [22].

B. Formant Frequencies

Speeches by different persons are distinguished by the short term spectral characteristics and temporal variations of the spectral characteristics[17]. This frequency is known as the fundamental frequency or pitch. In speech science and phonetics, formant is also used to mean an acoustic resonance of the human vocal tract[9]. Linear prediction based spectrum calculation is used for formant extraction[11]. Formants are local maxim in the speech spectrum. For a normal speech about five formant frequencies can be extracted. Mainly F1 and F2 frequencies represent the variations in pronunciation of phones. These frequencies are used to analyze the changes in the articulation problems of the speech[3]. These two frequencies is needed for disambiguating vowels. These two formants determine the quality of vowels in terms of the open/close and front/back dimensions. During speech the phonemes the formant frequency value changes. The time course of these changes in vowel formant frequencies are referred to as 'formant transitions'[8]. The frequency spectrum of normal speech has more formal transitions /spectral slope due to frequent formant transitions. The dysarthric speech spectrum has negligible formant transition frequency resulting in a small spectral slope[3].

C. System Architecture

The aim of the system is to enhance dysarthric speech by inducing above mentioned formant frequencies transitions. Formant frequencies increases the spectral slope of dysarthric speech, resulting in enhancement. First formant(F1) and second formant(F2) of dysarthric speech are used mainly focused as these formants alone are used to disambiguate the vowels[5]. Architecture of the system is shown in the figure.1.

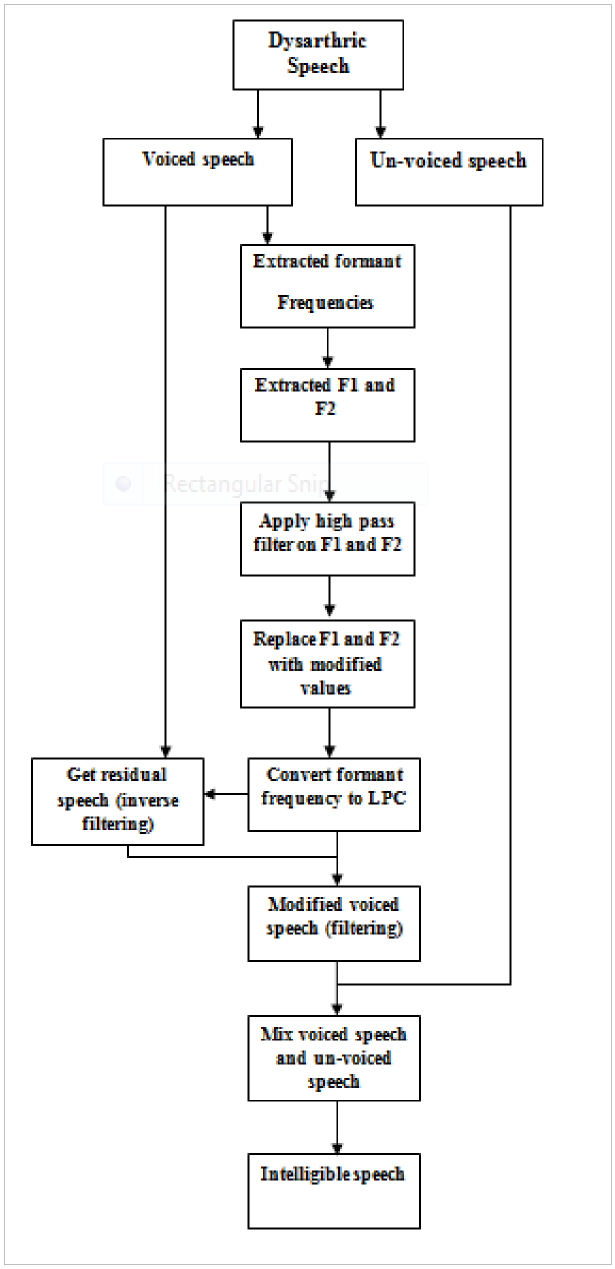

Figure 1. Formant Trajectory Refinement

Dysarthric speech enhancement using formant trajectory refinement system initially extracts the voiced speech and unvoiced speech separately from the dysarthric speech data using 'Praat-Vocal tool kit'[21]. Voiced speech in this context is the data with information i.e. vowels along with consonants. Unvoiced speech refers to fillers, noise, gap etc in a speech.
The voiced speech is used for the extraction of formant frequency. The LPC filter of 'Praat' tool is used for extracting the formant values from the extracted voiced data as per burg-algorithm[10]. The extracted format values are stored as object text file. The F1 and F2 formant frequency values alone of each frames of the object text file are extracted separately using Matlab[15]. These extracted frequency from each frames are then filtered using a 4-order high pass filter, thereby inducing slope. The changed frequency values after inducing slope are reflected on the object text files. These formant frequencies are then converted to LPC. After that inverse filtering is applied on the

LPC and original voiced sound to get the residual speech. Then filtering is done to the residual speech and LPC to get the modified voiced speech. Then voiced speech and unvoiced speech are mixed together to get the enhanced speech of dysarthric patient.

III. EXPERIMENTAL RESULTS

Objective quality measures are used to evaluate speech quality[14]. For evaluating speech, International Telecommunication Unit (ITU) provides standards known as quality measures. These quality measures are independent of speaker characteristics, hence known as objective quality measures. The objective measures are of two categories; single ended measurements and double ended measurements[18]. Single ended measures does not need reference signal, speech signal itself can be measured on some predefined properties. Double ended measures need reference speech signal for comparison. Dysarthric Speech Enhancement System uses ITU-T P.563 standard [19]. ITU-T P.563 is a single ended measure of speech quality which gives in subjective output ranging from 1 (min value) to 5 (max value) [1-bad, 2-poor, 3-fair, 4-good and 5-excellent][12]; hence referred to as mean opinion score (MOS). Average MOS of each directory of mild dysarthria is given in the table-1.

A good objective quality measure should have a high correlation with many different subjective measurements of human.

Table -1 Single Ended Measure(P.563)

Directory	MOS- Before Enhancement	MOS- After Enhancement
BB	1.8162	2.5357
FB	2.1928	2.7518
LL	1.7208	2.6166
MH	1.8870	2.1024

Double ended measurement used in dysarthric speech enhancement system is PESQ (Perceptual Evaluation of Speech Quality). PESQ uses ITU-T standard P.862, which measures audible speech quality with a reference signal, hence known as intrusive algorithms [11]. P.862 based on psycho-acoustics considerations and are trained on subjective databases to represent human perception like frequency selectivity, nonlinear response of human hearing, masking effects, critical band concept, and loudness. The resultant objective score values are in between 1 (min value) to 5 (max value) like MOS [20].

Composite measures [11] are obtained by combining basic objective measures to form a new objective measure. This measure rates the quality of the enhanced speech on three different counts. They are signal distortion(C_{sig}), background noise intrusiveness(C_{bak}) and overall quality(C_{ovl}) [12].

1) The speech signal alone using a five-point scale of signal distortion (C_{sig}) [1-Very unnatural, 2-Fairly unnatural, 3-Somewhat natural, 4-Fairly natural and 5-Very natural]
2) The background noise alone using a five-point scale of background intrusiveness (C_{bak}) [1-very intrusive, 2-somewhat intrusive, 3-Noticeable but not intrusive, 4- Somewhat noticeable and 5-Not noticeable]
3) The overall quality (C_{ovl}) [1-bad, 2-poor, 3-fair, 4-good and 5-excellent].

The higher values of Csig, Cbak and Covl represent lower signal distortion, lower background intrusiveness, and higher quality of speech, respectively. These values are obtained by linearly combining the existing objective measures by the following relations [11]:

$$Csig = 3.093 - 1.029LLR + 0.603PESQ - 0.009WSS$$
$$Cbak = 1.634 + 0.478PESQ - 0.007WSS + 0.063segSNR$$
$$Covl = 1.594 + 0.805PESQ - 0.512LLR - 0.007WSS$$

where LLR, PESQ, WSS, and segSN R represent the log likelihood ratio, perceptual evaluation of speech quality, weighted slope spectral distance and segmental SNR, respectively [12]. The resultant objective score values are in between 1 to 5 like MOS.

Table -2 Double Ended Measure(P.862 and Composite Measure)

Directory	Csig	Cbak	Covl	PESQ(ITU-T P.862)
BB	4.453	3.591	4.523	4.482
FB	3.846	3.531	4.218	4.482
MH	4.276	3.619	4.424	4.453
LL	3.819	4.000	4.214	4.493

IV.CONCLUSION

In this paper we are narrating a new methodology for the dysarthric speech enhancement. This methodology is mainly focused on medium intelligible improvement of dysarthric speech with mild disability, independent of speaker characteristic. Dysarthric speech enhancement system make use of the formant frequency value refinement which helps in preventing the de-coarticulation problems mainly by concentrating on the vowels and consonant intelligibility. Therefore, a more sophisticated formant trajectory model is likely to improve the intelligibility of dysarthric speech thereby supporting in dysarthric rehabilitation.

A highly intelligible, but much more error-prone methodology can be built by coupling a speech recognizer with this dysarthric speech enhancement system therby functioning as an intelligent interpreter.

REFERENCES

[1] Alexander Kain, Xiaochuan Niu, John-Paul Hosom, Qi Maio and Jan van Santen, "Formant Re-synthesis of Dysarthric Speechs", in 5 th ISCA Speech Synthesis Workshop, 2004.

[2] Barbara Tomik, Jery Krupinski, Lidia Glodzik-Sobanska, Maria Bala-Slodowska, Wieslaw Wszolek, Monika Kusiak and Anna Lechwacka "Acoustic Analysis of Dysarthria profile in ALS patients" in Journal of the Neurological Sciences pp 169-3542, 1999.

[3] Bjrn Lindblom, Diana Krull, Lena Hartelius Ellika Schalling " Formant transitions in normal and disordered speech: An acoustic measure of articulatory dynamics, " in Proceedings, FONETIK , Dept.of Linguistics, Stockholm University 2009.

[4] Frank Rudzics "Acoustic transformations to improve the intelligibility of dysarthric speech," in SLPAT '11 Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies pp.11–21.

[5] Gary Weismer and Jill Wilderrnuth "Fomant Trajectory Characteristics in Persons with Parkinson, Cerebellar,and Upper Motor Neuron Disease," in Waisman Center and Dept. Communicative Disorders, University of Wisconsin-Madison, Madison, WI 53705

[6] Hesham Tolba and Ahmed S. EL Torgoman " Towards the Improvement of Automatic Recognition of Dysarthric Speech, " in Iwanami Lecture Series: Language Sciences, Iwanami Shoten Publishers,1997.

[7] Lalitha V, Prema P and Lazar Mathew " A Kepstrum Based Approach for Enhancement of Dysarthric Speech, " in 2010 3rd International Congress on Image and Signal Processing (CISP)2010.

[8] Lingyun Gu , John G. Harris , Rahul Shrivastav and Christine Sapienza" Disordered Speech Evaluation using Objective Quality Measures, " in Proceedings of the ICASSP, 2005.

[9] Michael Robb and Michael Blomgren, " Analysis of F2 Transitionsin the Speech of Stutters and Nonstutters, " in J. Fluency Disord. 22 (1997), by Elsevier Science Inc., pp. 1-16, 1997.

[10] Paul Boersma David Weenink, Praat Doing Phonetics with Computer, www.praat.org.Version 5, 2008.

[11] P. Krishnamoorthy, " An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms, " in IETE TECHNICAL REVIEW , VOL 28 , ISSUE 4, 2011.

[12] P C. Loizou, Speech Enhancement: Theory and Practice, 1st edition Boca Raton, FL: CRC; 2007.

[13] Prema P, Vidhya S and Lazar Mathew T " Prototype for Dysarthric Speech Enhancement, " in ISSN: 0976-8882 E-ISSN: 0976-8890, Vol. 2, Issue 2, pp-50-53,2011.

[14] Richard Hummel " Objective Estimation of Dysarthric Speech Intelligibility, " in Queens University Kingston, Ontario, Canada, 2011.

[15] The Mathwoks Inc. - Matlab the Language of Technical Computing, www.Mathworks.com, Version 2011b

[16] Saranya. M, P. Vijayalakshmi and Nagarajan Thangavelu " Improving the Intelligibility of Dysarthric Speech by Modifying System parameters, Retaining Speakers Identity, " in ICRTIT ISBN: 978-1-4673-1601-9/12/31.00 IEEE 2012.

[17] Tamas Bohm and Geza Nemeth, "Algorithm for formant tracking, modification and synthesis" in VOLUME LXII, pp.1116, 2007.

[18] Tiago H. Falk, Student Member, IEEE, and Wai-Yip Chan, " Single-Ended Speech Quality Measurement Using Machine Learning Methods," in IEEE Transactions on Audio, Speech, and Language Processing, VOL. 14, NO. 6, November 2006.

[19] www.itu.int/rec/T-REC-P.563/" visited March 2013.

[20] www.itu.int/rec/T-REC-P.862/" visited March 2013.

[21] www.praatvocaltoolkit.com/" visited May 2013.

[22] Xavier Menndez-Pidal, James B. Polikoff, Shirley M. Peters, Jennie E. Leonzio and H. T.Bunnell "The Nemours Database of Dysarthric Speech," in Applied Science Engineering Laboratories (ASEL),A.I. duPont Institute, P.O. Box 269, Wilmington, DE 19899, USA