INVESTIGATING ARTICULATING DISORDERS USING EMPIRICAL MODE DECOMPOSITION

G. Georgoulas*, V. Georgopoulos** and C. D. Stylios***

* School of Electrical & Computer Eng. Georgia Institute. of Technology, Atlanta, GA, USA
**Dept. of Speech and Language Therapy, TEI of Patras, Koukouli Patras, Greece
***Dept. of Informatics and Communications Technology, TEI of Epirus, , Artas, Greece

ggeorgoulas@mail.gatech.edu, voula@teipat.gr, stylios@teiep.gr

Abstract: This paper presents a preliminary study of the applicability of a novel signal processing technique as a means to exact valuable information so that to diagnose the possible existence of a speech articulation disorder in a speaker. Articulation, in effect, is the specific and characteristic way that an individual produces the speech sounds. Emprirical Mode Decomposition and the Hilbert Huang transform is applied in an attempt to identify potential features to be used in an articulator disorder detector.

Introduction

Articulation refers to the production process of speech sounds in isolation or in words. The process describes the physiological movements involved in modifying the airflow, in the vocal tract above the larynx, for the production of the various speech sounds. In essence sounds, syllables, and words are formed when the vocal chords, tongue, jaw, teeth, lips, and palate change the stream of air that is produced by the respiratory system. Articulation is a complicated procedure that is often difficult to master. An articulation problem appears when a person produces sounds, syllables or words incorrectly so that listeners do not understand what is being said or they have to pay more attention to the way the words sound than to what he/she means. Most articulation errors fall into one of three categories: omissions, substitutions, or distortions.

In a typical substitution error, for example, a child may say /θ/ instead of /s/ in the Greek word /sela/ (saddle) so it would be heard as /θela/. Another case is the omission error where the second syllable of the word may be omitted leaving only /se/. These kinds of mistakes are systematic, which means that a child may only misarticulate a couple of sounds, but he/she does so in all words that contain those sounds. In many cases, this disorder results in unintelligible speech while in other cases the speech remains intelligible. This is a fact that depends on the frequency of the misarticulated sounds. In any of these cases, the articulation disorder constitutes a serious communication problem for the patient that has to be diagnosed so that to be solved through training .

From the clinical practice and experience [1], a few of the most common substitution articulation errors that Greek children make are shown in Table 1.

Table 1: Some of the common articulation errors in Greek

Target sound	Produced sound
ɾ	/γ/
/s/	/ʃ/, /θ/, or /ç/
/v/	/f/
/ð/	/θ/

The area of speech processing is an active and interesting area of signal processing and much work has been done for event detection in speech signals [2]-[3]. In this research work we propose the use of a novel signal processing technique to analyze the speech signal in an atempt to find a characteristic footptint for each one of the aforementioned articulation disorders.

Most real life processes are inheritably nonlinear and nonstationary. As a result, using techniques that assume linearity and stationarity even though they are build upon solid mathematical background can be suboptimal, misleading or even have completely no connection to the physical system that they are supposed to "model". Empirical Mode Decomposition (EMD) and the Hilbert Huang transform, introduced in [4] came to fill this gap between theory and real life. EMD lacks rigorous mathematical analysis and it decomposes the signal into a collection of Intrinsic Mode Functions (IMFs), where an IMF represents a simple oscillatory function with a number of conditions that have to be satisfied. The well behaved Hilbert transforms of the IMFs give an alternative approach to time-frequency decomposition which results from the traditional short time Fourier transform and the most recently developed wavelet transform [4].

In this research work we investigate the use of EMD and Hilbert Huang transform as a means to analyze the speech signal in order provide a more suitable representation that can be eventually combined with an advanced learning paradigm from the field of pattern recognition for the discrimination of articulation disorders.

This paper is organized: in the following section the data set used for analysis is described and then a brief description of the EMD algorithm and then Hilbert-Huang spectrum is presenting. The results of the application of EMD and Hilbert-Huang transform to normally and misarticulated phonemes are discussed and finally, conclusions and future directions are included.

Materials and methods

Using a computer with a sound card and an ordinary microphone, samples of 16-bit precision at a sampling rate of 44.1 KHz were collected from 16 children ages 6-8 whose mother tongue was Greek. All children were asked to produce the pseudoword /asa/. Speech therapists were used as experts to evaluate and categorize the articulation of children. Of the 16 children 4 had normal production of the pseudoword, and 12 produced articulation errors of which 4 were substitution of /s/ with /ʃ/, 4 were substitution of /s/ with /θ/ and 4 were substitution of /s/ with /ç/.

Empirical Mode Decomposition and the Hilbert-Huang spectrum

EMD is an algorithm that decomposes a signal into a fine set of oscillatory components (IMFs). These functions are symmetric with respect to a local zero mean and have the same number of zero crossings and extrema. The method for computing these functions was originally introduced by Huang et al. [4] and is implemented through the following steps [6]:

1. Identify all minima and maxima of the given signal ($x(t)$)
2. Create an upper ($e_{max}(t)$) and a lower ($e_{min}(t)$) envelope interpolating between successive maxima and minima respectively (usually via cubic interpolation)
3. Calculate the running mean
$$m(t)=\frac{e_{\min}(t)+e_{\max}(t)}{2}$$
4. Subtract the mean from the signal to extract the detail $d(t)=x(t)-m(t)$
5. Repeat the whole process replacing $x(t)$ with $m(t)$ until the final residual is a monotonic function (or a user specific number of IMFs has been extracted – application dependant)..

In practice, step 5 may not produce a valid IMF. As a result the sifting needs to take place which implies the iteration of steps 1 to 4 upon the detail d(t) until this fulfils the criteria of an IMF. Therefore the original signal x(t) is eventually decomposed into a sum of IMFs plus a residual

$$x(t)=\sum_i IMF_i(t)+r(t)$$

as it is shown in Figure.1.

Following the EMD the Hilbert transform can be applied to each IMF separately and the instantaneous frequency can be calculated as the derivative of the phase function. After performing the Hilbert transform to ach IMF the original signal can be expressed as the real part, RP, in the following form

$$x(t)=RP\left(\sum_j a_j(t)e^{i\theta_j(t)}\right)=RP\left(\sum_j a_j(t)e^{i\int\omega_j(t)dt}\right)$$

The above equation gives both the amplitude and the frequency of each component as a function of time. This time-frequency distribution of the amplitude is called the Hilbert-Huang spectrum ($H(\omega,t)$). Integrating over time we get the marginal spectrum.

$$h(\omega)=\int_0^T H(\omega,t)\,dt$$

The marginal spectrum offers a measure of total amplitude (or energy) contribution from each frequency value.

The frequency in either $H(\omega,t)$ and $h(\omega)$, has a totally different meaning from the Fourier spectral analysis [4]. While in the classical Fourier representation the existence of energy at frequency, ω, means a component of a sine or a cosine wave persisted through the whole time span, in the case of the of the marginal spectrum the existence of energy at the frequency ω, means only that in the whole time span, there is a higher likelihood for such a wave to have appeared locally.

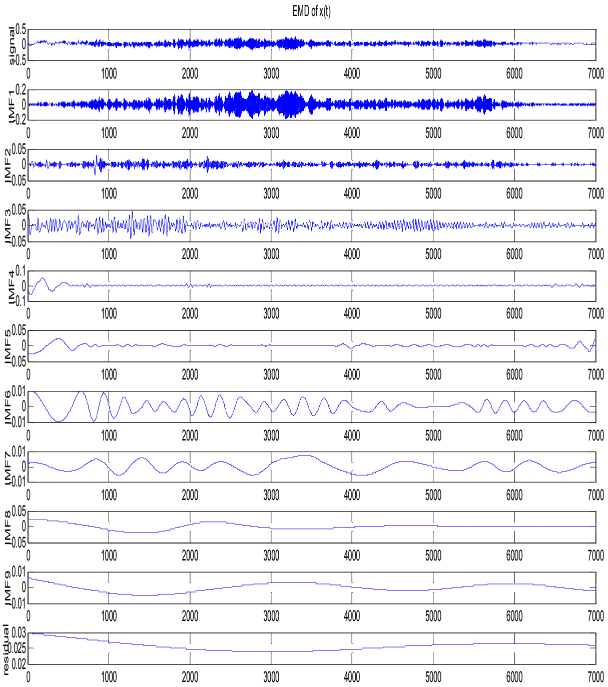

Figure 1: Application of the EMD algorithm to phoneme /s/ produced by a normal speaker

Results

The implementation of the EMD has been performed using the freely available MATLAB toolbox by Rilling et al. [6],[7] along with the TFTB toolbox developed by the same group [8].

In our first attempt to investigate the utility of EMD in the analysis of phonemes we focused on the analysis of the marginal spectrum h(ω) of the first 4 IMFs. As it can be seen in Figures 2-5 the 3 types of disorders give rise to a somewhat different "energy concentration" as it is depicted in terms of the marginal spectrum.

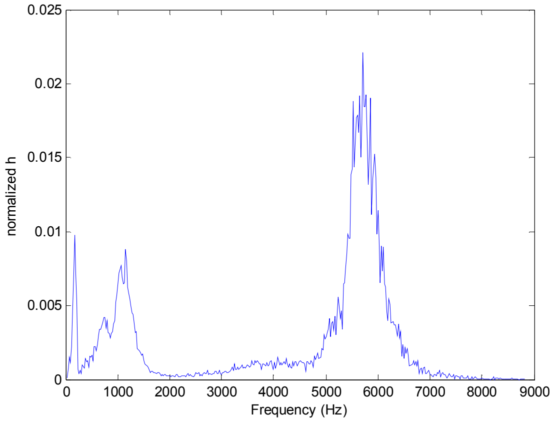

Figure 2: Normalized of marginal *h*(*ω*) of phoneme /s/

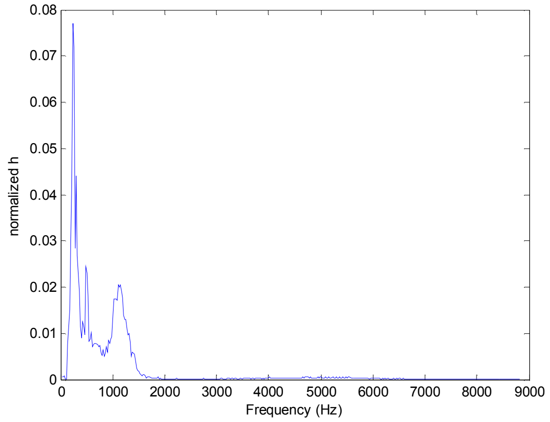

Figure 3: Normalized of marginal *h*(*ω*) of phoneme /θ/

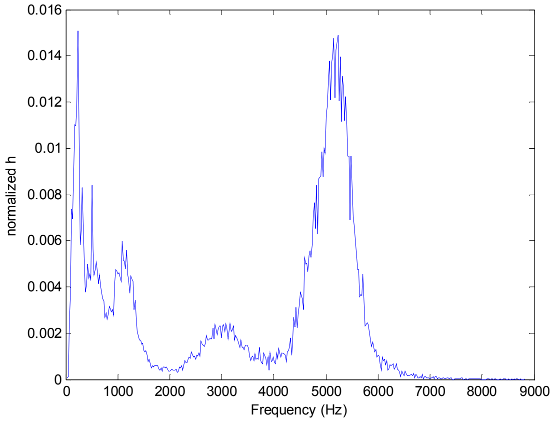

Figure 4: Normalized of marginal *h*(*ω*) of phoneme /ç/

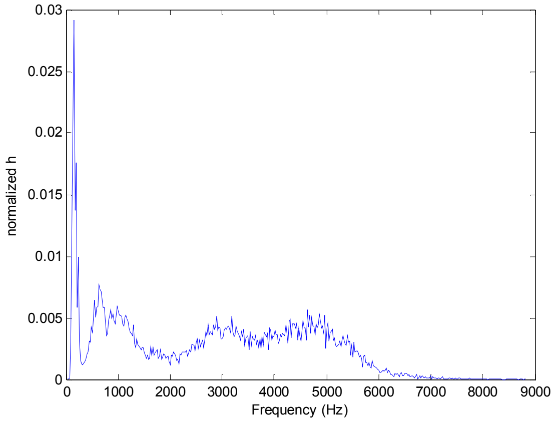

Figure 5: Normalized of marginal *h*(*ω*) of phoneme /ʃ/

Figures 2-5 come from 4 different individuals. Merging together the marginal spectrums of the different individuals we come up with Figure 6.

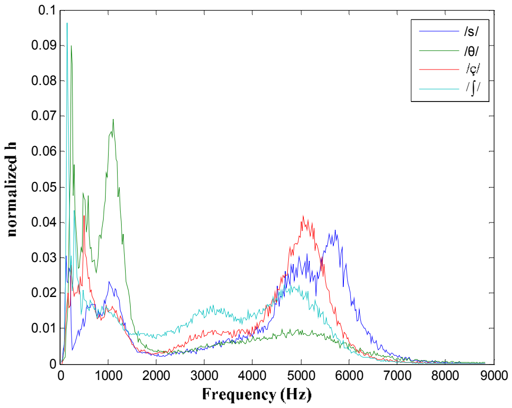

Figure 6. Averaged normalized marginal spectrum across all subjects

As it can be seen the normally pronounced phoneme /s/ gives rise to higher frequencies followed by the erroneously produced phoneme /ç/. Phonemes /θ/ and /ʃ/ seem to have most of their concentrated in lower frequencies.

Conclusions

The proposed method to analyze correctly pronounced and misarticulated phonemes seems to give promising results according to this preliminary study, but there are still some issues that have to be considered. First of all the whole analysis was restricted on a qualitative analysis of the results. A more rigorous quantitative analysis has to be performed that will also involve the extraction of specific features to be used in a second stage responsible for the automatic classification of the different phonemes.

Furthermore, only the marginal spectrum of the first 4 IMFs was employed in this analysis. The time evolution of the speech signal was not explicitly taken into account. In future work we will try to exploit this specific capability provided by the Hilbert-Huang transform.

Finally even though the proposed method is promising, it still has to be tested using a larger data set with words, pseudowords, and continuous speech before safer conclusions can be drawn

Acknowledgements

This research work was supported by the joint research project “Intelligent System for Medical Data Analysis and Evaluation Based On Computational Intelligence Techniques” under the Greek-Czech bilateral cooperation agreement, funded by Greek GSRT and Czech MEYS.

References

1. V. C. Georgopoulos, G. A. Malandraki, and C. D. Stylios, (2001) A computer based speech therapy system for articulation disorders, in *Proc. 4th Int. Conf. Neural Networks and Expert Systems in Medicine and Healthcare (NNESMED)*, Milos Island, Greece, Pages. 223-230
2. B. Yegnanarayana and R. N. J. Veldhuis (1998) Extraction of vocal-tract system characteristics from speech signals, *IEEE Trans. on Speech and Audio Processing*, Volume. 6, Number. 4, Pages. 313-327.
3. H. Kawahara, Y. Atake, and P. Zolfaghari (2000) Accurate Vocal Event Detection Method based on a Fixed-point to Weighted Average Group Delay, in *Proc. of ICSLP'2000*, Beijing, China, Volume. IV, Pages. 664-667
4. N. E. Huang, Z. Shen, S. R. Long, M. L.Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu (1998) The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society London A*, Volume 454, Pages 903–995
5. I. Daubechies (1992), *Ten Lectures On Wavelets*. Philadelphia: Siam
6. G. Rilling, P. Flandrin and P. Goncalves (2003) On Empirical Mode Decomposition and its Algorithms, *in Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, NSIP03*, Grado, Italy
7. http://perso.ens-lyon.fr/patrick.flandrin/emd.html
8. http://tftb.nongnu.org/