Chapter 8
Automatic Determination of Pauses in Speech for Classification of Stuttering Disorder

João Paulo Teixeira
Polytechnic Institute of Bragança, Portugal

Maria Goreti Fernandes
Polytechnic Institute of Bragança, Portugal

Rita Alexandra Costa
Polytechnic Institute of Bragança, Portugal

ABSTRACT

An algorithm to automatically identify segments of silence or speech is presented. The algorithm was developed to measure the silence periods in spontaneous and read speech. These silence periods are one of the parameters used to know the degree of severity of stuttered speech. For this purpose the three longer disfluent events (pauses or other disfluent events) and also the percentage of silence are useful. The algorithm is based on the evaluation of the energy and the zero crossing rate of the signal compared to the threshold values previously determined in silence. One experiment with eight subjects is described using the Stuttering Severity Instrument for Children and Adults – SSI and the percentage of silence in speech. It was concluded that the percentage of silence is good enough to separate stuttered from the normal speech but alone is not capable of measuring the degree of severity of the stuttered speech.

DOI: 10.4018/978-1-5225-1724-5.ch008

INTRODUCTION

Speech is one of the most fundamental and complex cognitive human acts. The normal speech is the final product of a complex network of linguistic, cognitive and sensorimotor processes. Its production requires the coordinated activation of distinct muscle systems and the vocal tract (Juste et al., 2012; McClean & Tasko, 2004).

Considering the analysis of the speech signal, there are three different states of the speech: silence, unvoiced speech and voiced speech. In the silence state, no speech is produced and the muscles within the vocal folds are relaxed. In the unvoiced state, the folds are closer together and tenser than in the silence state, allowing a turbulence to be generated at the folds themselves. In the voiced state, active and passive contractions of the chest and abdominal wall generate a subglottic pressure that exceeds the closure force of the adducted vocal folds. The transglottic air pressure differential produces an airflow that is modulated by the vocal folds to produce a time-varying longitudinal air pressure wave. This pressure wave is changed by the vocal tract to create the sound we hear as the normal human voice (Plant & Younger, 2000).

For a given idiom, there are a set of phonemes that characterize the language. These phonemes can be divided into vowels and consonants. The vowels group contains the oral, nasal and semi-vowels or glides. The consonants are divided in plosive, liquid, fricatives and vibrant. The plosive vowels are composed by the occlusive part (almost or completely occlusion of sound) and by the plosive part generally followed by one vowel but sometimes followed by other consonant. Anyhow, each of these consonants can be voiced or unvoiced. The voiced sounds are produced with the vibration of the vocal cord and the unvoiced sounds are produced without vibration of the vocal cords and with the glottis open. The voiced sounds generally have low frequency energy and in opposition unvoiced sounds has higher frequency energy. The frequency of vibration of the vocal cords is known as the fundamental frequency (F0), which is controlled by the states of tension and length of the vocal cords. Greater tension and length correspond to higher frequency tones, (Seeley, Stephens & Tate, 2006).

Speech disorders are human disabilities that affect millions of people worldwide and are usually treated with behavioral therapy (Barnes et al., 2016). It is estimated that 40 million Americans have a communication disorder (Ancelle, 2015). The study and evaluation of human speech disorders may lead to a wider array of treatment options and provide key insights into the genetic and neural underpinnings of human speech. Developmental stuttering is the principal disorder of fluency (Ancelle, 2015). This speech disorder is characterized by frequent occurrences of repetitions or prolongations of syllables, words, and sounds, as well as involuntary hesitations or pauses that disrupt the rhythmic flow of speech (Wieland et al., 2015). In addition to the changes in the rhythm of speech, the stuttering is commonly accompanied by body movements, such as tremors, spasms of oro-facial and laryngeal muscles, and also abnormal involuntary movements (ticks) (Mulligan et al., 2003; Riva-Posse et al., 2008; Rogić et al., 2016). Stuttering onset usually begins between the ages of two and five years, when children begin to form simple sentences (Vanhoutte et al., 2016). Recent studies have indicated that the incidence of stuttering is approximately 5%, however the majority of affected children (about 80%) recovers during the puberty (Wieland et al., 2015; Rogić et al., 2016; Ancelle, 2015). It is estimated that 1-2% of world adult population continues to suffer from severe stuttering (often called Persistent Developmental Stuttering) (Prado-Velasco & Fernández-Perunchena, 2011).

In addition to involuntary speech disruptions, stuttering can affect nearly all aspects of a person's life. Different definitions of stuttering refer to specific emotions such as fear, anxiety, embarrassment and irritation (Adriaensens, Beyers & Struyf, 2015). The importance of this subject have motivated the development of several studies. Van Lieshout, Hulstijn & Peters (2004) stated that the motor abilities (speech production system) of stutterers are limited in motion. And Peters, Hulstijn & Van Lieshout (2000) affirmed that these motor abilities could be represented in a continuous way from the normal to a deviant behavior. The cited limited motor abilities (Van Lieshout, Hulstijn & Peters 2004) means that differences between the speech of stutterers and speech from fluent persons are not permanent and happen mainly when the stutter motor system (stutterer) control is destabilized. This means that the fluent speech of persons who stutter may already have differences to the speech of fluent persons.

Another feature that can influence the stuttering level of a stutterer is the speech rate. Hirsch (2007) stated that the increase of the speech rate is a destabilizing factor of the stutter motor system. He also stated that in face of fast speech rate the stutterers do not show the 'undershoot phenomenon'. The undershoot of vowel targets in fluent persons underlies a reduction of the movement amplitude of the motor system, so as to make up for the added cost required by the accelerated speech rate.

Other authors Yaruss (1999) and Sawyer, Chon & Ambrose (2009) stated that the speech production is also perturbed by the phonological complexity. That perturbation of the motor system control was found for Spanish and English adult stutterers (Howell & Au-Yeung, 2007; Howell, Au-Yeung, Yaruss & Eldridge, 2006). A sequence of clusters at the beginning of words also increases the disfluency in adult stutterers (Howell, Au-Yeung & Sackin, 2000). Regarding these references it can also be stated that the motor behavior of stutterers may also show differences from one language to another.

Typically, the frequency of stuttering is measured by counting the number of stutters that are judged to occur in a sample and reporting this as a proportion of the total amount of speech that occurred. In other words, the percentage of syllables spoken that were stuttered (Costello & Ingham, 1984). The preferred denominator is number of syllables spoken (as opposed to words) because it more accurately depicts the amount of speech produced by controlling for variation in word length and the amount of stuttering produced by permitting counts of multiple stuttering per word. Furthermore, when data are collected in real time (i.e., concurrent with ongoing speech), the listener can more easily recognize individual syllables, the pulses of which are relatively salient, compared to words, which require more cognitive processing by the listener.

There are several counting processes of syllables and words, in order to classify the degree of stuttering. In most cases, the counting is made in real time through the listener and following a live or audio/audio-video presentation of the speech sample. This process is adequate as a sensitive indicator of stuttering frequency (Martin, Kuhl, & Haroldson, 1972; Onslow, Andrews, & Lincoln, 1994; Onslow, Costa, & Rue, 1990; Reed, & Godden, 1997; Ryan, & Ryan, 1995).

Inherent to the use of counts of stutters is the thorny issue of how stuttering is defined by the listener. Different definitions such as behavioral, perceptual, and even speaker-determined, abound in the literature (Conture, 1990; Curlee, 1981; MacDonald & Martin, 1973; Martin & Haroldson, 1981; Perkins, 1990; Wingate, 1964; Yairi, 1997; Young, 1975).

Currently, there is no research that irrefutably proves one definition more valid than the others. Therefore, it is important that researchers clearly describe the way in which stuttering was identified and recorded by listeners and include information regarding those listeners, their training, their experience, and, of course, the reliability of their counts. Readers can then judge for themselves the quality of the stuttering identification method and hence the quality of the reported findings.

Table 1. Scores and degrees of severity for each of the subjects

Score									Severity Degree
#	Gender	Age	Frequency			Duration of the 3 Longer Events	Behaviors Associated	Final Score	
			Spontaneous Speech	Reading Speech	Total				
1	M	23	9	5	14	8	10	32	Serious
2	M	12	9	9	18	14	14	46	Very Serious
3	M	22	6	4	10	8	10	28	Moderate
4	F	18	5	4	9	10	6	25	Moderate
5	F	21	2	0	2	4	0	6	Not Disfluent
6	F	25	0	2	2	4	0	6	Not Disfluent

(Teixeira, Fernandes & Costa, 2012; Teixeira, Fernandes & Costa, 2013)

In the literature, there are some specific classification instruments to evaluate the severity of stuttering speech. Riley's Stuttering Severity Instrument is one of the most common scales (Riley, 1972). Through this tool, a previous study was developed to classify the degree of severity of stuttering in a group of six subjects (Teixeira, Fernandes & Costa, 2012; Teixeira, Fernandes & Costa, 2013). The typology followed was similar to that of author Wendell Johnson (1963), since it showed episodes of stutter that fit more with the scale of measurement of the degree of severity of disfluency. These tests were implemented using spontaneous and reading speech. The behavior of the subject were observed and registered during the speech record. The evidence allowed obtaining the Frequency, which is the number of disfluent syllables divided by the total number of spoken syllables, expressed as a percentage, and the Duration, which is the average of the three longer disfluent events over both tests (spontaneous speech and reading speech). The values of these two parameters were converted to a score, using the standard scales proposed by Silva (2009). By using the Matlab® software, an algorithm was developed to automatically count the parameter for further evaluation of the degree of severity of stutter, as well as obtaining the means of the three longer disfluent events.

Finally, statements were added to the scores of the three stages of evaluation (Frequency, Duration of the three longer disfluent events and Behaviors associated with). The final value was converted into percentages for comparison with limit tabled values, allowing the classification of the degree of severity of disfluency in very low, low, moderate, severe and very severe degrees. Table 1 shows the results obtained in the above mentioned study. It should be mentioned that the 6 subjects were also used in the study of the percentage of silence. The first four subjects are known as persons with stuttering.

SCALE: STUTTERING SEVERITY INSTRUMENT FOR CHILDREN AND ADULTS

The scale "Stuttering Severity Instrument for Children and Adults" (SSI-3 and SSI-4) proposed by Riley (1972, 2009) is a common behavioral assessment tool for measuring stuttering severity among three age groups; preschool, school age and adults. This instrument provides an estimate of stuttering sever-

ity using three types of parameters: frequency (percentage of the stuttered syllables), duration (average duration of the three longest stutters) and observations of physical concomitants made at the time of the recording (i.e. distracting sounds, facial grimaces, head movements and movements of the extremities) (Todd et al., 2014). The SSI-3 has been investigated by the scientific community and classified as being a reliable scale (Ansari et al., 2010). For the application of this scale it is necessary to perform the test using spontaneous and reading speech as well as an observation and further evaluation of behaviors along the two records.

The symptoms that are considered as stutters are identical in SSI-3 and SSI-4 and are described in detail in the SSI-3 manual (Riley, 1994). All versions of the SSI scores arise from measures made on both the speech sample and observations of physical concomitants. There are standard scales that convert the percentage of stuttered syllables and duration measures into scores that are combined with raw physical concomitant scores to give the total overall SSI score (Todd et al., 2014). The scale classifies the subject in one of the five groups according to the observed behaviors (0 = none, 1 = not visible unless you're watching, 2 = barely noticeable to a casual observer, 3 = distracting, 4 = Very distracting, 5 = Severe and painful to look). The record of the two parameters over the speech will allow reaching scores of observed behaviors, using a normative scale. This score is then used to classify the degree of disfluency. Finally add up the scores of the three evaluation parameters (Frequency, Duration of the three longer diffluent events and Associated Behaviors (or Physical Concomitants)) to give a final value that is converted to a percentage, which enables, when compared to standard values, the classification of the disfluency in a severity scale in Very Low, Low, Moderate, Severe and Very Severe.

The SSI-4 includes a computer program that automates the assessment of stuttering severity (not available in SSI-3). However, according to the authors, Howell *et al.*, (2011), the program has not been assessed for reliability and validity. Also, results with the program have not been compared against the methods for obtaining stuttering severity recommended in SSI-3 (Howell *et al.*, 2011).

Some flexibility exists in the procedures that can be used to obtain speech samples. Provision is made in the SSI instrument for the assessments to be made in clinical, home, or laboratory settings.

The minimum required sample length is specified. The minimum sample length given in Riley reduced the minimum sample length from 200 to 150 syllables for SSI-4 (Todd et al., 2014).

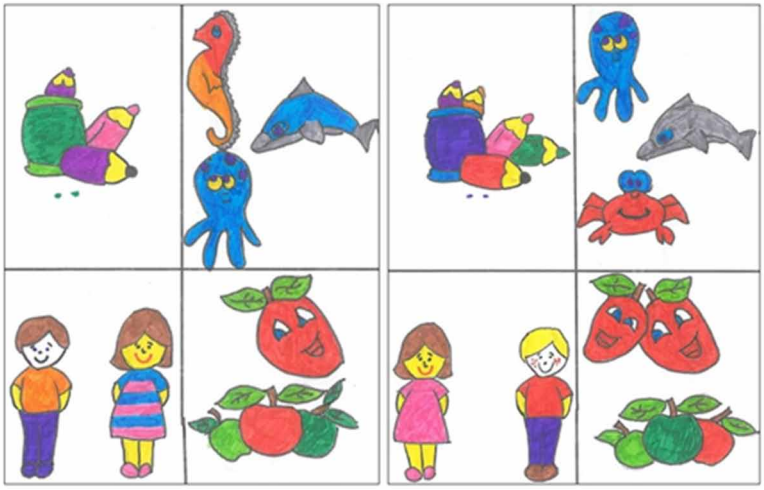
METHODOLOGY

This section describes the methodology of the experiment to measure the stuttering severity of a group of persons in order to explore the described scale. Next section will describe the implementation of the algorithm developed to measure the duration of pauses in the speech.

Participants

This research included a total of 8 subjects of which 6 had already participated in the previous study cited in Table 1. Participants of this research were 4 male adults, 3 female adults and 1 male child who volunteered to take part in the present study. All subjects gave their informed consent prior to participation in the study. For the underaged the permission was obtained from their legal representatives. None of the study participants had presented history of head trauma, learning disabilities, dyslexia, psychiatric conditions and use of any medication. Only the child was in speech therapy. Data were acquired on eight

Figure 1. Example of images used to tease the spontaneous conversational speech



native speakers of Portuguese Language who reported having normal hearing. Four subjects (between the ages of 11 and 23 years) reported a history of stuttering since childhood, and the other 4 subjects (between the ages of 22 and 26 years) had a normal speech production.

Speech Fluency Evaluation

The stuttering severity levels were determined from videotape recordings of spontaneous speech and reading tasks, using the Stuttering Severity Index or SSI (Riley 1972). The present analysis focuses exclusively on two tasks. The first task aimed to analyze a conversational speech (about 15 minutes of duration) between a stutterer participant and a person with normal speech, without visual contact between them. To establish communication, several similar images were given to the two participants in order to discuss the images between them (Figure 1). In the second task, the participants were invited to read a text containing 318 words and 1559 fluent syllables for analysis and comparison of the results.

Speech Signal Record

To ensure that the information obtained is reliable, it is necessary that the recording of the signal is made under adequate conditions and using the adequate equipment.

There are several processes that enable the acquisition of speech signal for later analysis. One which is commonly used is a process where the acquisition of the acoustic signal produced by the speaker is made in a soundproof room, using a unidirectional or omnidirectional microphone. This signal is amplified by a pre-amplifier with a linear frequency response and is stored on a magnetic tape with good quality, being later or immediately held for their conversion to digital signal using an anti-aliasing filter. The storage of the signal is then done in digital format, (McAulay & Quatieri, 1986; Teixeira, 2013). It is important to use a sampling frequency high enough to guarantee a bandwidth to ensure the required quality for understanding the speech. Anyhow, once the requirements for understanding the speech do not need the higher frequencies, the sampling frequency of 11025 Hz is enough.

In the present work, the data acquisition was carried out in an appropriate and quiet room, trying to be close to a soundproof room. Participants spoke into two *Sennheiser Best e840* unidirectional microphones (frequency ranges = 40-18000 Hz) positioned directly in front of their mouths. Each trial was recorded and exported into the Praat software (Boersma & Weenink, 2009) generating .wav files. This software allowed a segmentation of the speech signal for a later use. The Record parameters were mono sound, a 16-bit resolution and a sampling frequency of 11025 Hz. Finally, MATLAB (Mathworks, Inc.) was used to develop and implement the pause measurement algorithm in each speech signal of the participants.

AUTOMATIC DETERMINATION OF PAUSES IN SPEECH

The measurement can be made using some speech signal analysis tools like Praat or SFS (Speech File System). These tools allow to represent and ear the segments of speech and label them, this allows to manually identify the silent segments in speech and measure them. Anyhow all the signals need to be manually labeled, that is a time consuming task. Alternatively an algorithm was developed in order to label and identify automatically the speech signal into silent or speech segments.

The silent segments in regular speech may occur between sentences or paragraphs, the longer segments of silence, occasionally between words, shorter segments, and also during the stop part of the stop consonants (<k>, <p>, <t>, <g>, <d> and <b>), being very short time pauses (between 40 and 70 ms in average for read speech in European Portuguese, according to Teixeira et al., 2001). Additionally, some interjections may occur during reading but mainly during spontaneous speech. Some of the interjection contains silent parts of speech. In stuttered speech several additional silent parts of speech are inserted. All these different types of silent parts of speech will be measured by the algorithm.

For the automatic determination of silence or pauses in speech some tools were used, among them the Moving Average (MA), the Moving Energy (ME) and the Zero Crossing Rate (ZCR). These tools perform processing only in the temporal domain.

The algorithm was originally developed by Teixeira (2013) for the classification of the signal in the zones of silence, voiced speech and unvoiced speech. The algorithm is based on the three mentioned tools (moving average, the moving energy and the zero crossing rate) and in one decision area. This decision is based on the result of two vectors obtained by the moving energy and zero crossing rate.

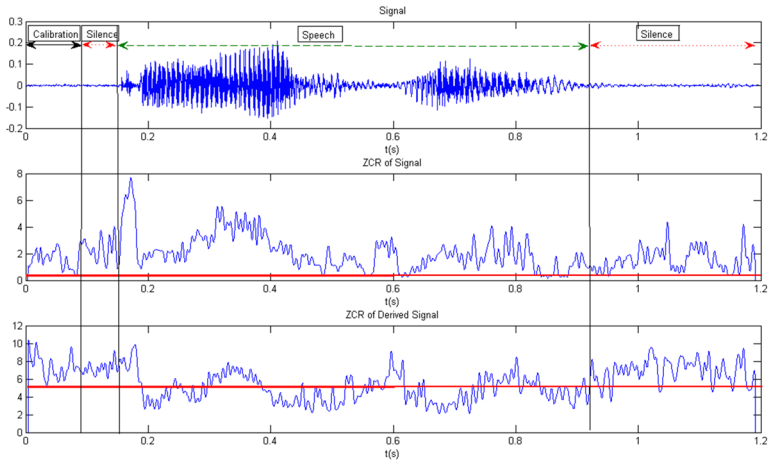
Zero Crossing Rate

Zero Crossing Rate gives the number of times the signal crosses the zero line during a defined length of signal. It can be expressed by Eq. 1 and 2.

$$ZCR(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} \left| sign(x(m)) - sign(x(m-1)) \right| \times w_{N+1}(m) \quad (1)$$

where:

Figure 2. Application of the zero crossing rate. a) Speech signal corresponding to the Portuguese word "casa" (home); b) Zero crossing rate applied directly to the signal; c) Zero crossing rate applied to the derived signal



$$sign(x(m)) = \begin{cases} 1 & \text{if } x(m) \geq 0 \\ -1 & \text{if } x(m) < 0 \end{cases} \quad (2)$$

w is a window with length N+1 and x is the input signal. Different types of windows can be used. In this algorithm a rectangular window was used.

The zero crossing rate is applied to the derived signal and not directly to the original signal because the original signal often was an offset. This offset even if it is small usually leads to the noisy level of the signal above or below zero eliminating the zero crossing.

The derivative consists of a simple difference applied by the following equation:

$$d(n) = x(n) - x(n-1) \quad (3)$$

The application of the zero crossing rate to the derivative of the signal (d) consists essentially of the determination of the rate of the high and low peaks. This parameter is high inside the parts of silent speech, actually noise speech, low in voiced speech, and also high in unvoiced speech.

The advantage of using the derived signal instead of the original signal to apply the zero crossing rate is demonstrated in Figure 2. In this figure the signal corresponds to the Portuguese word "casa".

The first 150 ms correspond to a silent part of the signal, meaning high ZCR expected, then the plosive consonant <k> corresponds to unvoiced speech and therefore also high ZCR although slightly lower than in the silent part of the signal. After the consonant (about the instants of 190 ms) it follows the <a> vowel, that is a voiced sound and therefore with low ZCR. Then the next sound is the voiced fricative consonant <z>. It is expected to have low energy compared to voiced sounds, but higher energy than silence segments, and higher ZCR than in voiced segments. After the consonant comes the closed voiced vowel <a> with low ZCR and higher energy. Finally, the signal ends with a silent part corresponding to high ZCR. The upper part of Figure 2 represents the speech signal and the calibration, silence and

speech segments are identified. The middle part of the figure represents the application of the ZCR to the original signal and lower part the application of the ZCR to the derived signal. The initial and final part of the signal has a very low offset, not visible in the figure, but enough to deviate the noise signal from zero line and reduce the ZCR, as it can be observed in the middle part of Fig. 2. It can be seen that the application of the zero crossing rate to the original signal do not gives high values due to this slight offset, but the application to the derived signal already detects this high values correspondent to the silent segments.

The ZCR was applied under a window of 20 samples and with a displacement of 1 sample in both cases presented in Figure 2.

Moving Average and Moving Energy

The moving average function implements Eq. 4. This function has the ability to smooth the signal. The signal is more or less smoothed according to the length of the window N+1. For this purpose the Hanning window was used for w. The indices n do not need to be of unitary space, it can be highly spaced to have higher efficiency.

$$MA(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} x(m).w_{N+1}(m) \tag{4}$$

The moving energy function implements Eq. 5. It determines the energy of the signal and is more or less smoothed according to the length of the window N+1.

$$ME(n) = \frac{1}{N+1} \sum_{m=n-N/2}^{n+N/2} \left[x(m).w_{N+1}(m)\right]^2 \tag{5}$$

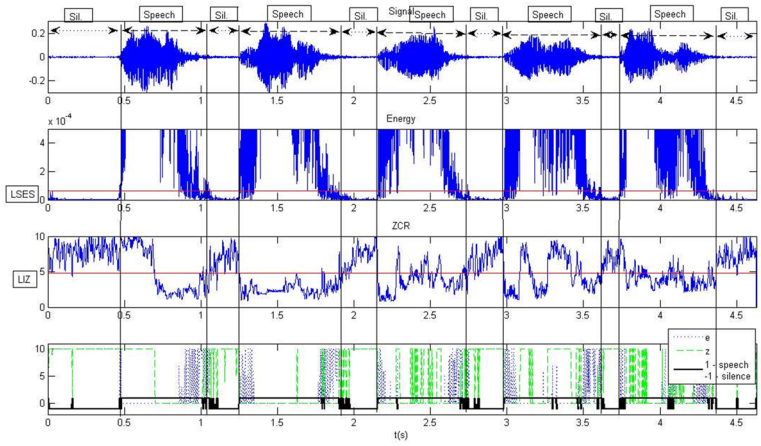
Algorithm

After reading the signal the algorithm calls the moving energy function with a window of length 20 and spaced with 10 samples allowing smoothing of the signal. This new signal, ME, will have a length 10 times lower than the original. Then, the original signal is derived to produce the d signal, and the zero crossing rate function is applied to the d signal with a window of 20 samples and a spacing of 10 (both cases correspond to a superposition of 50% of the segments). The output signal (ZCR) is smoothed out by applying the moving average function with a window length of 100 samples and unitary spacing. Also this signal will have a length ten times less than the original.

The beginning of the signal (about 100 ms) is used to calibrate the level of silence in the environment. During this period no speech should be produced and a silence is expected in order to calibrate the level of noise namely its energy and ZCR. This calibration will serve to define the threshold values for the energy and the zero crossing rate for the silence.

In Figure 3, secured silence is in the first instants of time, at least 100 ms. This beginning of the signal with silence is a noisy signal with low energy and a high zero crossing rate. This part of silent signal is

Figure 3. Classification of the signal with paused read speech corresponding to "Eu moro na minha casa" (I live in my house) according to the field of decision established



used to determine the upper limit of energy (LSES) and the lower limit of zero crossing rate (LIZ) in silent part of the signal. In the second signal of Figure 3 there is a visible threshold line that relates to the maximum level of energy corresponding to the silence, called LSES (upper limit of the energy signal), and this is the maximum of the moving energy. In turn, the minimum level of zero-crossing rate after being smoothed is also established, called LIZ (Lower Limit of ZCR), which is visible by a threshold line in the third signal of Figure 3.

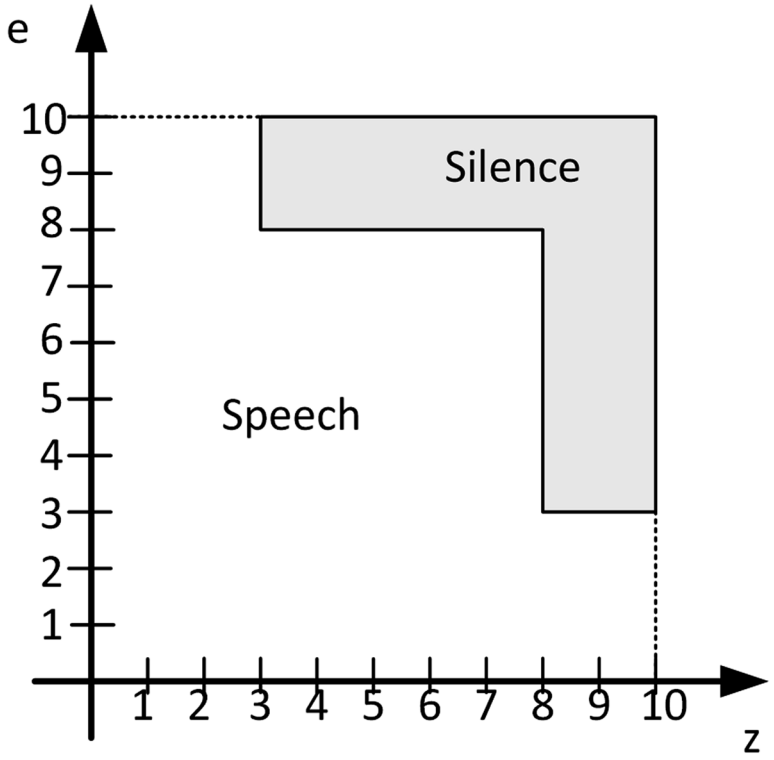
Continuing the follow-up of the algorithm, the variables e and z were declared for counting the number of elements below the LSES threshold and above the threshold LIZ respectively under a window of length 10. This window with 10 elements corresponds to a 100 samples of the original signal which corresponds to 9 ms long. Therefore, every 9 ms window there will be a decision about the type of signal. High values of e represent low-energy, high values of z imply high ZCR. Based on the variables e and z a decision is made to classify the original signal Silence or Speech.

On the basis of variables e and z a "matrix of decision" was established, having the z variable in the abscissa, between 0 and 10 and the e variable in the ordinate axes, also between 0 and 10. In the matrix of decision the areas were settled for the silence and speech, as can be seen in Figure 4.

In Figure 4 the areas of classification of speech signal can be observed, where the silence corresponds to: $z \geq 8 \text{ ^ } e \geq 3$ or $z \geq 3 \text{ ^ } e \geq 8$. The first area corresponds to high ZCR and relatively low energy, the second area corresponds to relatively high ZCR and very low energy.

The second and third signals of Figure 3 present the ME (moving energy) and ZCR. The fourth signal presents the e and z and the final decision of silence/speech for each segment of 9 ms. It can be seen that the decision here correctly assumed along of the all signal in Figure 3. For this example the measured time with speech was 4.62 s, the time with silence was 1.56 s, given a percentage of silence of 33.7%, measured according to Eq. 6. It should be mentioned that this signal was read in a paused speech to show the decision between silence and speech segments and does not correspond to a stuttered speech.

Figure 4. "Matrix of decision" for classification of speech signals based on e and z variables



Program Code

Next lines present the main algorithm to make the decision of silence or speech for each segment of the signal. Each segment has duration of 9 ms.

```
Detrend filter applied to the speech signal
Moving Energy of the signal
Determination of the threshold of energy in silent speech (LSES)
Signal is derived
The zero crossing rate (ZCR) is applied to the derived signal
A moving average is applied to the ZCR
Determination of the threshold of ZCR in silent speech (LIZ)
Cycle 1, % for i=1:end - segments of 9 ms
    z=0; % number of elements above LIZ
    e=0; % number of elements below LSES
    Cycle 2 (for j=1:10)
        if ZCR(j)>LIZ,
           z=z+1;
        end
        if ME(j)<LSES,
```

```
            e=e+1;
        end
    end cycle 2
    if ((e>=8 & z>=3) or (e>=3 & z>=8)),
      decision(i)= silence
      else
      decision(i)= Speech
      end
end of cycle 1
```

ANALYSIS OF RESULTS

With the application of the algorithm to the recorded signals it was possible to make the collection of the parts of silence and parts of speech of all the segments. These results were useful to determine the percentage of silence in each of the 8 subjects. The percentage of silence was obtained based on the following equation:

$$\%\ of\ silence = \frac{Time\ of\ Silence}{Time\ of\ Speech} \times 100 \tag{6}$$

Table 2 shows the values obtained for the percentage of silence in the disfluent and control subjects. The first 6 subjects correspond to the subjects presented already in Table 1, using the same number of subject. To obtain these results, 5 tracks of the signal of each of the subjects were selected which ensured equal conditions of measurement. The segments under analysis were edited leaving an initial portion of speech signal that would guarantee at least 100 ms of silence for the determination of the LSES and LIZ, and a start and end time that corresponded to the beginning and end of the speech, respectively.

As shown in Table 2, the subjects with disfluencies present higher percentage values of silence, with an average of 26.3%. In this group the percentage of silence varies from 20.0 to 41.8%. Subject number 2, presenting 41.8% of silence is the subject among the 4 disfluent belonging to the sample under study that will present a disfluency more accentuated as recorded in Table 1, on the previous study (Teixeira, Fernandes & Costa, 2012; Teixeira, Fernandes & Costa, 2013). Based on results presented in Tables 1 and 2 it can be alleged that it is not a viable classification of degree of disfluency solely based on the percentages of silence. For instance, the subject 1, ranked with a degree of severity serious has the lowest percentage of silence considering the subjects 3 and 4, classified with a degree of moderate severity. This can be explained by the fact that there is a number of parameters, in addition to the silence that influence the classification of the degree of severity.

The percentage of silence in the control group (not disfluent) is minor than the ones of the disfluent subjects. For this group the percentage of silence varies from 6.2% to 16.2%, and has an average of 11.0%. All the controls have lower values than any disfluent.

Table 2. Results obtained for the percentage of silence in all subjects

	Subject	% of Silence
Disfluent	1	20.0
	2	41.8
	3	23.2
	4	20.2
	Average	26.3
Control subjects	5	6.2
	6	16.2
	7	6.8
	8	14.8
	Average	11.0

One aspect that may alter results is the emotional state of the subject, because when it is under the pressure of an evaluation it may become nervous, causing a possible increase of disfluent moments throughout the speech. In addition, to being under pressure for an assessment, the fact of having to run the tests with strange people nearby can also cause changes in the occurrence of disfluent moments.

CONCLUSION

In this work an algorithm to automatically measure the silences in speech was developed. The algorithm is based on the determination of the energy and zero crossing rate for small windows of the signal. The zero crossing rate is applied to the derivate signal to recover from the eventual offset. The threshold values LSES and LIZ, the upper limit of energy and lower limit of zero crossing rate, respectively, in the silence part of speech is determined. Then the values of e and z are determined in a window with 10 samples. After that the e and z are used to decide if the segment with 10 samples, corresponding to 9 ms, is silence or speech. This algorithm was applied to determine the percentage of silence in spontaneous disfluent speech and in a normal spontaneous speech. This spontaneous speech was recorded putting two subjects discussing similar but different pictures and separately recording each subject.

The algorithm allowed the determination of the percentage of silence of each subject.

The analysis of the percentage of silence shows that this parameter clearly separates the subjects with and without disfluency. It is possible to conclude that individuals with disfluency show percentages of silence much higher compared with individuals without any kind of disturbance in speech.

Six of the eight subjects were in a previous study which was used to measure their degree of severity of the disfluency. The percentage of silence was compared to the previously determined degree of severity and it is possible to conclude that this measure alone is not able to determine the degree of severity of the disfluent speech.

REFERENCES

Adriaensens, S., Beyers, W., & Struyf, E. (2015). Impact of stuttering severity on adolescents' domain-specific and general self-esteem through cognitive and emotional mediating processes. *Journal of Communication Disorders*, *58*, 43–57. doi:10.1016/j.jcomdis.2015.10.003 PMID:26484722

Ancelle, J. A. G. (2015). Assistive Technologies at the Edge of Language and Speech Science for Children with Communication Disorders: VocalIDTM, Free SpeechTM, and SmartPalateTM. In N. R. Silton (Ed.), *Recent Advances in Assistive Technologies to Support Children with Developmental Disorders* (pp. 255–277). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8395-2.ch012

Ansari, H., Bakhtiar, M., Ghanadzade, M., Packman, A., & Seifpanahi, S. (2010). Investigation of the reliability of the SSI-3 for preschool Persian-speaking children who stutter. *Journal of Fluency Disorders*, *35*(2), 87–91. doi:10.1016/j.jfludis.2010.02.003 PMID:20609330

Barnes, T. D., Wozniak, D. F., Gutierrez, J., Han, T. U., Drayna, D., & Holy, T. E. (2016). A Mutation Associated with Stuttering Alters Mouse Pup Ultrasonic Vocalizations. *Current Biology*, *26*(8), 1009–1018. doi:10.1016/j.cub.2016.02.068 PMID:27151663

Boersma, P., & Weenink, D. (2009). *Praat Manual: doing phonetics by computer. 5.1.17.* [Computer program]. Available at: http://www.fon.hum.uva.nl/praat/download_win.html

Conture, E. (1990). *Stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

Costello, J. M., & Ingham, R. J. (1984). Assessment strategies for stuttering. In R. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions* (pp. 303–333). San Diego, CA: College-Hill Press.

Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research*, *24*(4), 595–560. doi:10.1044/jshr.2404.595 PMID:7035743

Hirsch, F. (2007). *Le bégaiement Perturbation de l'organisation temporelle de la parole et conséquences spectrales*. (Ph.D dissertation). Marc Bloch Univ., Strasbourg.

Howell, P., & Au-Yeung, J. (2007). Phonetic complexity and stuttering in Spanish. *Clinical Linguistics & Phonetics*, *21*(2), 111–127. doi:10.1080/02699200600709511 PMID:17364620

Howell, P., Au-Yeung, J., & Sackin, S. (2000). Internal Structure of Content Words Leading To Lifespan Differences in Phonological Difficulty in Stuttering. *Journal of Fluency Disorders*, *25*(1), 1–20. doi:10.1016/S0094-730X(99)00025-X PMID:18259599

Howell, P., Au-Yeung, J., Yaruss, S., & Eldridge, K. (2006). Phonetic difficulty and stuttering in English. *Clinical Linguistics & Phonetics*, *20*(9), 703–716. doi:10.1080/02699200500390990 PMID:17342878

Howell, P., Soukup-Ascencao, T., Davis, S., & Rusbridge, S. (2011). Comparison of alternative methods for obtaining severity scores of the speech of people who stutter. *Clinical Linguistics & Phonetics*, *25*(5), 368–378. doi:10.3109/02699206.2010.538955 PMID:21434809

Johnson, W. (1963). *Diagnostic Methods in Speech Pathology*. New York: Herper & Row.

Juste, F. S., Rondon, S., Sassi, F. C., Ritto, A. P., Colalto, C. A., & Andrade, C. R. (2012). Acoustic analyses of diadochokinesis in fluent and stuttering children. *Clinics (Sao Paulo)*, *67*(5), 409–414. doi:10.6061/clinics/2012(05)01 PMID:22666781

MacDonald, J., & Martin, R. (1973). Stuttering and disfluency as two reliable and unambiguous response classes. *Journal of Speech and Hearing Research*, *17*(4), 691–699. doi:10.1044/jshr.1604.691 PMID:4783809

Martin, R., & Haroldson, S. (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research*, *24*(1), 59–63. doi:10.1044/jshr.2401.59 PMID:7253630

Martin, R. R., Kuhl, P., & Haroldson, S. K. (1972). An experimental treatment with two preschool stuttering children. *Journal of Speech and Hearing Research*, *15*(4), 743–752. doi:10.1044/jshr.1504.743 PMID:4652394

McAulay, R., & Quatieri, T. (1986). Speech Analysis/Synthesis Based on a Sinusoidal Representation. *IEEE Transactions on Acoustics Speech, and Signal Processing, 34*(4).

McClean, M. D., & Tasko, S. M. (2004). Correlation of orofacial speeds with voice acoustic measures in the fluent speech of persons who stutter. *Experimental Brain Research*, *159*(3), 310–318. doi:10.1007/s00221-004-1952-8 PMID:15248043

Mulligan, H. F., Anderson, T. J., Jones, R. D., Williams, M. J., & Donaldson, I. M. (2003). Tics and developmental stuttering. *Parkinsonism & Related Disorders*, *9*(5), 281–289. doi:10.1016/S1353-8020(03)00002-6 PMID:12781595

Onslow, M., Andrews, C., & Lincoln, M. (1994). A control/experimental trial of operant treatment for early stuttering. *Journal of Speech and Hearing Research*, *37*(6), 1244–1259. doi:10.1044/jshr.3706.1244 PMID:7877284

Onslow, M., Costa, L., & Rue, S. (1990). Direct early intervention with stuttering: Some preliminary data. *The Journal of Speech and Hearing Disorders*, *55*(3), 405–416. doi:10.1044/jshd.5503.405 PMID:2381182

Perkins, W. H. (1990). What is stuttering? *The Journal of Speech and Hearing Disorders*, *55*(3), 370–382. doi:10.1044/jshd.5503.370 PMID:2199728

Peters, H., Hulstijn, W., & Van Lieshout, P. (2000). Recent Developments in Speech Motor Research into Stuttering. *Folia Phoniatrica et Logopaedica*, *52*(1-3), 103–119. doi:10.1159/000021518 PMID:10474010

Plant, R. L., & Younger, R. M. (2000). The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech. *Journal of Voice*, *14*(2), 170–177. doi:10.1016/S0892-1997(00)80024-7 PMID:10875568

Prado-Velasco, M., & Fernández-Peruchena, C. (2011). An Advanced Concept of Altered Auditory Feedback as a Prosthesis-Therapy for Stuttering Founded on a Non-Speech Etiologic Paradigm. In J. Pereira (Ed.), *Handbook of Research on Personal Autonomy Technologies and Disability Informatics* (pp. 76–118). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-206-0.ch006

Reed, C., & Godden, A. (1997). An experimental treatment using verbal punishment with two preschool stutterers. *Journal of Fluency Disorders*, *2*(3), 225–233. doi:10.1016/0094-730X(77)90026-2

Riley, G. (1972). A Stuttering Severity Instrument for Children and Adults. *The Journal of Speech and Hearing Disorders*, *37*(3), 314–322. doi:10.1044/jshd.3703.314 PMID:5057250

Riva-Posse, P., Busto-Marolt, L., Schteinschnaider, A., Martinez-Echenique, L., Cammarota, A., & Merello, M. (2008). Phenomenology of abnormal movements in stuttering. *Parkinsonism & Related Disorders*, *14*(5), 415–419. doi:10.1016/j.parkreldis.2007.11.006 PMID:18316236

Rogić Vidaković, M., Jerković, A., Jurić, T., Vujović, I., Šoda, J., Erceg, N., & Dogaš, Z. et al. (2016). Neurophysiologic markers of primary motor cortex for laryngeal muscles and premotor cortex in caudal opercular part of inferior frontal gyrus investigated in motor speech disorder: A navigated transcranial magnetic stimulation (TMS) study. *Cognitive Processing*, 1–14. PMID:27130564

Ryan, B. P., & Ryan, B. V. K. (1995). Programmed stuttering treatment for children: Comparison of two establishment programs through transfer, maintenance, and follow-up. *Journal of Speech and Hearing Research*, *38*(1), 61–75. doi:10.1044/jshr.3801.61 PMID:7731220

Sawyer, J., Chon, H., & Ambrose, N. (2009). Influences of Rate, Length, and Complexity on Speech Disfluency in a Single Speech Sample in Preschool Children Who Stutter. *Journal of Fluency Disorders*, *33*(3), 220–240. doi:10.1016/j.jfludis.2008.06.003 PMID:18762063

Seeley, R., Stephens, T., & Tate, P. (2006). *Anatomy and Physiology (7th ed.)*. Mcgraw-Hill.

Silva, S. (2009). *Classificação do grau de disfluência com e sem o uso de feedback acústico modificado em adolescentes e adultos gagos portugueses*. (Unpublished undergraduate dissertation). University Fernando Pessoa, Portugal.

Teixeira, J. P. (2013). *Análise e Síntese da Fala - Modelação Paramétrica de Sinais Para Sistemas TTS*. Editorial Académica Espanhola.

Teixeira, J. P., Fernandes, M. G., & Costa, R. A. (2012). Measure and Comparison of Speech Pause Duration in Subjects with Disfluency Speech. *Procedia Technology*, *5*, 812–819. doi:10.1016/j.protcy.2012.09.090

Teixeira, J. P., Fernandes, M. G., & Costa, R. A. (2013). Pause Duration of Disfluent Speech. *International Journal of Reliable and Quality E-Healthcare*, *2*(3), 62–73. doi:10.4018/ijrqeh.2013070105

Teixeira, J. P., Freitas, D., Braga, D., Barros, M. J., & Latsch, V. (2001). *Phonetic Events from the Labeling the European Portuguese Database for Speech Synthesis, FEUP/IPB-DB*. Eurospecch.

Todd, H., Mirawwdeli, A., Costelloe, S, Cavenagh, P., Davis, S., & Howell, P. (2014). Scores on Riley's stuttering Severity Instrument Versions Three and Fur for samples os dofferent length and for different types of speech material. *Informa Healthcare, 28*(12), 912-926.

Van Lieshout, P., Hulstijn, W., & Peters, H. (2004). Searching for the weak link in the speech production chain of people who stutter: a motor skill approach. In B. Maassen, R. Kent, P. H. van Lieshout, & W. Hulstijn (Eds.), Speech motor control in normal and disordered speech. Oxford University Press.

Vanhoutte, S., Cosyns, M., van Mierlo, P., Batens, K., Corthals, P., De Letter, M., & Santens, P. et al. (2016). When will a stuttering moment occur? The determining role of speech motor preparation. *Neuropsychologia*, *86*, 93–102. doi:10.1016/j.neuropsychologia.2016.04.018 PMID:27106391

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S. E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language*, *144*, 26–34. doi:10.1016/j.bandl.2015.03.008 PMID:25880903

Wingate, M. (1964). A standard definition of stuttering. *The Journal of Speech and Hearing Disorders*, *29*(4), 484–489. doi:10.1044/jshd.2904.484 PMID:14257050

Yairi, E. (1997). Disfluency characteristics of childhood stuttering. In R. F. Curlee & G. M. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (2nd ed.; pp. 49–78). Boston: Allyn & Bacon.

Yaruss, J. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research: JSLHR*, *42*(2), 329–344. doi:10.1044/jslhr.4202.329 PMID:10229450

Young, M. A. (1975). Onset, prevalence, and recovery from stuttering. *The Journal of Speech and Hearing Disorders*, *40*(1), 49–58. doi:10.1044/jshd.4001.49 PMID:1123928

KEY TERMS AND DEFINITIONS

Matrix of Decision: Matrix used to decide if the signal segment is silence or speech. It is based in 2 parameters related with the zero crossing rate and energy of the signal.

Moving Average: A signal with the result of the application of a moving average window along the signal. It is useful to smooth the original signal.

Moving Energy: A signal with the result of the application of a moving energy window along a signal. It gives the energy of the signal over the time.

Praat Software: Open source software to perform speech signal analysis. Useful also to phonetic studies (http://www.fon.hum.uva.nl/praat).

SSI: Stuttering Severity Instrument for Children and Adults is a common behavioral assessment tool for measuring stuttering severity.

Stuttering Disorder: Speech stuttering disorder characterized by frequent occurrences of repetitions or prolongations of syllables, words, and sounds, as well as involuntary hesitations or pauses that disrupt the rhythmic flow of speech.

Zero Crossing Rate: Gives the number of times the signal crosses the zero line during a defined period of time.