Time-frequency detection of stridence in fricatives and affricates

Slobodan Jovicic[a], Silvana Punisic[b] and Zoran Saric[b]

[a]School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serby
[b]Institute for Experimental Phonetics and Speech Pathology, Gospodar Jovanova 35, 11000 Belgrade, Serby
jovicic@etf.bg.ac.yu

As wheezes in abnormal breath sounds observed in patients with obstructive pulmonary diseases, the stridence in voice is manifested as excessively sharp, conspicuous, usually habitual hiss that is especially distinct with whispering. This paper reviews the articulator and acoustics features of stridence in fricatives and affricates, and presents an algorithm for detection of stridence. Detection of stridence was based on: time-frequency representation by FFT power spectra, time-frequency representation by AR-Burg power spectra, and power trajectory of signals in characteristic frequency bands. Many features are extracted from this analysis and served as input to the nonlinear classifier that decides about stridence in voice and the level of pathology. The algorithm was tested with speech database of normal and pathology voices. The results of automatic stridence detection showed high level of coincidence with the judgment of speech therapists.

1 Introduction

Disorders of voice, resonance, or articulation can affect the quality and clarity of expressive communication, making speech unpleasant to the listener, difficult to understand or even unintelligible. These communication disorders can have a negative impact on the patient's social interactions, interpersonal relationships, job opportunities, and certainly have a significant effect on the quality of life. Although the speech-language pathologists have expertise in assessing anatomic, physiologic and phonetics abnormalities in voice and speech production, they may have difficulty assessing the functional correlates to the abnormalities or distortions found.

Voice disorders can be recognized in the respiratory, phonatory and resonatory systems [1]. Mutual characteristic of these systems is that they function as unique voice production system which controls air flow from lung to mouth. Normally, this air flow generates a friction noise with specific spectral features. If the control of muscles tension is inadequate or some obstructions are present in the whole air flow tract the voice distortions appear with characteristic friction noise manifestations. Some specific air flow obstructions are recognized as *wheezing*, *stridor* or *stridence*.

Wheezing is abnormal breath sounds observed in patients with obstructive pulmonary diseases. It is manifested as high-pitched, musical, variable sounds with breathing, most prominently during expiration [2]. The sound is generated by air flowing through narrowed or irregular airways.

Stridor is specific form of wheezing (expiratory stridor) and is referred to upper airway obstructions [3]. It is the typically shorter, crowing sound which is often evident during inspiration (inspiratory stridor, due to the collapse of the soft tissues with negative pressure during inspiration).

Stridence in voice is manifested as obstructions in supraglottal part of the vocal tract. It is recognized as specific type of sigmatism with changed spectral picture of fricatives and affricates [4]. The stridence is speech sound produced by forcing air through a constricted passage between blade of tongue and hard palate characterized by a hissing sound. Special mode of stridence appears with diastema (space between two incisors).

Traditional assessment on voice disorders has relied on auditory perceptual judgment of voice quality by clinicians. Since the judgment is subjective in nature, instrumental measurements are needed to provide more evidence for systematic diagnosis of voice disorders. Examples of conventional instrumental measures include aerodynamic test, electromyography, laryngoscopy, radiology, MRI, and acoustic analysis [5]. Acoustic analysis appears promising computer approach in evaluation of voice disorders and interestingly, there exist many solutions in wheezing and stridor evaluation [6, 7, 8], but not for stridence. It is understandable because wheezing and stridor are located in subglottal airways and are more serious problems for patients.

On the other hand, stridence is more referred to the voice distortions at phonetic level and would be easily analyzed on speech signal. This paper presents acoustic analysis of stridence that appears in fricatives and affricates, and signal processing algorithm for detecting it, time-frequency stridence detector (TF-SD).

2 Characteristics of stridence sounds

Stridence is an acoustic occurrence that is generated in the mouth when certain conditions are achieved as to the position of the tongue in relation to the palate and teeth. Most often it is a particular form of constriction in the oral cavity through which an air current passes when certain voices are pronounced. This means that stridence appear simultaneously with the production of such voices.

In general, a constriction behaves as a source of a noise whose spectrum is modelled in the upper part of the vocal tract. A sound generated in this way, becomes a key element of voices such as fricatives and affricates. If the constriction takes on an irregular form then sounds with other content appear in addition to the regularly generated noise. From the acoustic aspect constriction in such conditions can have various forms such as: a short pipe, formed by a grove down the middle of the tongue, which is covered by the upper palate, or a resonant cavity, formed between the tip of the tongue and the upper incisors, etc. When the airflow reaches a certain speed while passing thru these acoustic elements, stationary waves and oscillations take place, whereas a tone of a certain frequency is generated, as with the flute, or a very strong resonant noise of narrowband character. These sounds are generated simultaneously with the pronunciation fricatives and affricates and they influence the changed spectral picture of friction and affrication. In other words, in addition to the normal friction and affrication, the described sounds which are called *stridence* appear. When perceived, stridence is experienced as an unpleasant, whistling, squeaky or coarse sound that influences the quality of the pronounced voice.

The formation of the described irregular constriction and the appearance of a stridence can also occur sporadically in normal speech, especially in transitional positions of the tongue in the oral cavity that are characteristic for co-articulated segments. The stridence is then quite short in nature and is not considered a pathological manifestation.

However, in speakers who have problems with motor control of the tongue (and perhaps some anatomical irregularities in the oral cavity) stridence occurs regularly in the production of certain voices and in that case is considered a pathological manifestation.

Fig.1 demonstrates a number of forms of stridence. Each example contains a signal waveform of the pronounced word, in which a fricative or affricate that contains a stridence is in the initial position, in addition to an appropriate narrowband spectrogram in which the effect of the stridence is better seen.

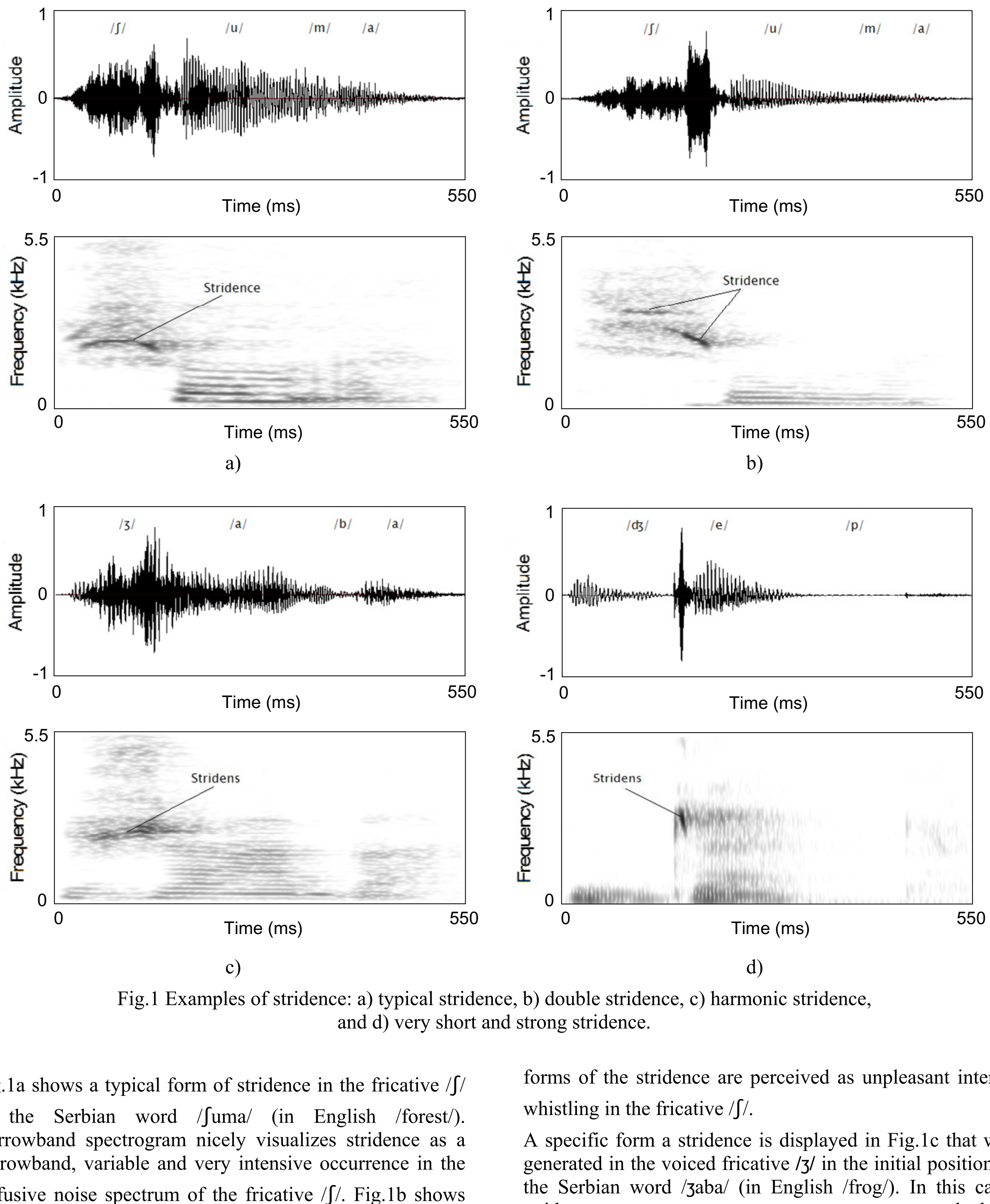

Fig.1 Examples of stridence: a) typical stridence, b) double stridence, c) harmonic stridence, and d) very short and strong stridence.

Fig.1a shows a typical form of stridence in the fricative /ʃ/ in the Serbian word /ʃuma/ (in English /forest/). Narrowband spectrogram nicely visualizes stridence as a narrowband, variable and very intensive occurrence in the diffusive noise spectrum of the fricative /ʃ/. Fig.1b shows the occurrence of a twofold stridence in the fricative /ʃ/ in the word /ʃuma/ that is spoken by another person. The realization of both stridence look are independent occurrences. In other words, a strident that is higher in the spectrum is completely stable in the spectrum, while a lower strident is very changeable in the spectrum. Both forms of the stridence are perceived as unpleasant intense whistling in the fricative /ʃ/.

A specific form a stridence is displayed in Fig.1c that was generated in the voiced fricative /ʒ/ in the initial position in the Serbian word /ʒaba/ (in English /frog/). In this case, stridence as a resonant occurrence encompassed three harmonics of the fundamental frequency of the voiced fricative /ʒ/ and significantly gained them in relation to the other part of the spectrum, together with the relevant part of the noised spectrum. Perceptually, this form of stridence significantly influences the quality of the fricative /ʒ/, when a „squeaky" voice is heard.

Fig.1d shows the stridence generated with the production of the voiced affricate /ʤ/ in the Serbian word /ʤep/ (in English /pocket/). The stridence is very short and very changeable in its frequency, and because of this it is shown with the aid of a broadband spectrogram. Perceptively it is experienced as a very piercing and unpleasant whistling.

The preceding analysis shows the following general characteristics of stridence:

- Stridence occurs in voices whose articulation requires constriction.
- Stridence appears simultaneously with the generation of friction and affrication.
- Stridence occurs more often in unvoiced fricatives and affricates because of the greater constriction.
- In the Serbian language, stridence most often occurs with the fricatives /ʃ/ and /ʒ/, within the frequency range of 2000 to 4000Hz, with the fricatives /s/, /z/ and affricate /ʦ/ with the frequency range from 6000 to 9000 Hz, with the affricates /ʧ/ and /ʤ/ in the freqency range from 3000 to 8000 Hz, and affricates /ʨ/ i /ʥ/ in the fruqency range of 3000 to 8000 Hz.
- Stridence is a resonant occurance with a spectral band of 25 to 250 Hz, which, in its intensity is about 20 dB above the envelope of the surrounding spectrum.
- The minimal duration of a stridence is approximately 10ms, whilst the maximum duration can be up to ¾ of the duration of the friction or affrication.
- There exists a strong correlation between the spectral and amplitude manifestations of the stridence.

3 Proposed TF-SD algorithm

3.1 Algorithm

The time-frequency algorithm for stridence detection (TF-SD) is the part of a screening system for assessment of distortion in expression of Serbian voices (Punišić et al., 2007). User of the system pronounce a word by word from database, system analyses speech signal and presents the result of voice distortions. One of the first steps in input speech signal pre-processing of this system is stimulus (word) segmentation to phoneme and subphoneme segments. In that way TF-SD algorithm has at his input fricative segments or affricate subsegments related to affrications.

Block diagram of TF-SD algorithm is shown in Fig.2. Speech signal is sampled with frequency f_s = 22050Hz. It is firstly processed by bandpass filter $BP(f_1,f_2)$, from the filter bank, that was selected according to the previous discussion about frequency distribution of fricatives and affricates. Filtered signal is divided on the overlapping frames $X(L)$, 512 samples wide, where L is number of frames. Further processing is performed in two steps. In the first step TF-SD algorithm determines potential position of the frame with stridence, and in the second step it has to decide if the stridence really exists there.

In the first processing step FFT is applied to each frame to obtain power spectrum density $S(f,L)$ and the sequence of signal power $P_s(L)$ in the range of f_1 to f_2. The local maximums in the sequence $P_s(L)$ are potential positions of the stridence. Hence, goal was to find out all the local maximums $P_{locmax}(L)$, then to calculate its strength $str(L)$ and finally, to select the frame L_{max} with maximal local strength by $L_{\max} = \arg\max_L(str(L))$. If the stridence is present, L_{max} is the segment in which the stridence is best heard.

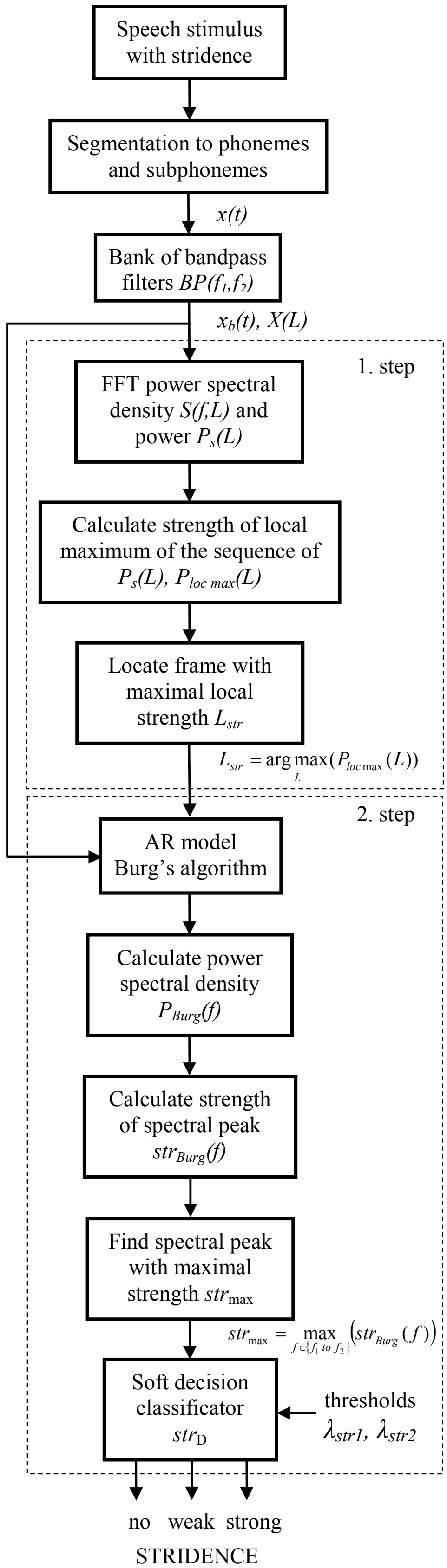

Fig.2 Block diagram of TF-SD algorithm.

When the potential position of the stridence is determined at L_{max}, it is needed to evaluate if the stridence really exists there. It is the next processing step.

In the spectrum domain the stridens is manifested as strong and narrow spectral line, Fig.1. This phenomenon was determined using AR model estimated by Burg's algorithm. AR model order 6 was estimated by Burg's algorithm on segment $X(L_{\max})$ and than the power spectrum $P_{Burg}(f)$ was calculated. Next step is calculation of local spectral maximum strength $str_{Burg}(f)$ through sliding spectral window. Generally, AR model order 6 has three spectral maxima and maximum of $str_{Burg}(f)$ is to be founds as $str_{\max} = \max\limits_{f \in \{f_1\ to\ f_2\}} \left(str_{Burg}(f) \right)$.

Finally, TF-SD algorithm performs detection and classification of stridence through soft decision by

$$str_D = \frac{1}{1 + \exp(-(str_{\max} - \lambda_{str}))},$$

where λ_{str} is experimentally determined threshold. The benefit of the soft decision compared to the hard decision is that expert can obtain information whether the stridence is week or strong. Because of that λ_{str} has two values λ_{str1} and λ_{str2} to makes distinction between three outputs: no stridence, weak stridence and strong stridence.

3.2 Testing

The preliminary results in testing of TF-SD algorithm are obtained on limited number of subjects with and without stridence. The group of 30 schoolchildren was tested, where 15 subjects had stridence and 15 subjects had not stridence. Distribution of stridence across fricatives and affricates was following: /ʃ/ - 8, /ʒ/ - 4, /ʧ/ - 2, and /ʤ/ - 1, and all stridences were strong.

Results of testing showed fully detection and discrimination of yes/no stridence and fully coincidence with the judgment of speech therapists. Certainly, a possible limitation of this testing is the small number of patients and no existence of weak stridence.

4 Conclusion

An efficient stridence detector, based on TF-SD algorithm, was presented in this paper that could be successfully used in diagnostic and rehabilitation practice due to its robustness and increased accuracy. This study was carried out in schoolchildren, and it would be of great interest to focus of future studies in stridence monitoring and testing of adult subjects and infants.

Acknowledgments

This work was partially supported by Ministry of Science of Republic Serbia under Grant number TP–6134B and Grant number 148028G.

References

[1] D. R. Boone, *"The voice and voice therapy",* (5th Ed.) Englewood Cliffs, Prentice Hall (1994)

[2] T.R. Fenton, H. Pasterkamp, A. Tal, V. Chernick, "Automated spectral characterization of wheezing in asthmatic children", *IEEE Transactions on biomedical engineering* BME-32 (1), 50-55 (1985)

[3] R.T. Cotton, J.S. Reilly, "Stridor and airway obstruction", in: C.D. Bluestone, S.E. Stool, M.A. Kenna (Eds.), *"Pediatric Otolaryngology",* W.B. Saunders Co, Toronto, pp. 1275–1276 (1996)

[4] J. Honova, P. Jindra, J. Pešák, "Analysis of articulation of fricative prealveolar sibilant "s" in control population", *Biomed. Papers* 147(2), 239–242 (2003)

[5] J. Hirschberg, T. Szende, P. Koltai, A. Illenyi, *"Cry, Stridor And Cough: Acoustic Imaging Analysis of the Pediatric Airway",* Plural Publishing (2008)

[6] J. E. Earis, B. M. G. Cheetham, "Current methods used for computerized respiratory sound analysis", *Eur. Respir. Rev.* 77(10), 586-590 (2000)

[7] B. S. Lin, B. S. Lin, H. D. Wu, F. C. Chong, S. J. Chen, "Wheeze recognition based on 2D bilateral filtering of spectrogram", *Biomedical Engineering Applications, Basis & Communications* Vol. 18, No. 3, 29-38 (2006)

[8] S. A. Taplidou, L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds", *Computers in Biology and Medicine* 37, 1073–1083 (2007)

[9] S. Punišić, S. Jovičić, Z. Šarić, "Speech distortion assessment: methodological and system approach", *Proc. of Int. Conf. SPECOM'07*, St. Petersburg, 355-360 (2007)