Int. J. Nonlinear Anal. Appl. 12 (2021) No. 2, 603-607
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.5115

The impact of subband pause noise on the sensitivity of detection methods

Entidhar Malik Hadi[a], Reham Ihssan Kamal[a], Ali M. Ahmed[a,*]

[a]*Al-Esraa University College, Baghdad, Iraq*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Through the use of subband analysis, the article presents a method for selecting pauses in speech communications. An additive blending of normally distributed noise was used in the sensitivity analysis. In various noise/signal respects, the probability of making a bad decision has been identified. The findings show that the proposed sub-band method is stable when subjected to additive noise effects.

Keywords: Speech Communications, sensibility, noise/signal, additive noise effects.

1. Introduction

One of the main stages of speech signal processing when solving problems such as speech recognition, speech compression, cleaning speech from noise, etc., it is the determination of areas of lack of speech - selection of pauses. The accuracy of determining the boundaries of pauses affects the efficiency of further stages of analysis and processing. In particular, the probability of correct recognition, speech sound quality and compression ratio. It is known that different speech sounds and noises in pauses have different energy distribution in the frequency domain. Taking these features into account allows you to determine the boundaries between speech sounds and pauses. Studies show that when implementing methods for selecting pauses, it is necessary to take into account several characteristics of the compared signal segments. The main test hypothesis is formulated as follows.

H_0 analyzed segment PC $\vec{x} = (x_1, x_2, \ldots, x_N)^T$ is generated by noise in the speech pause

$$\vec{x} = \vec{u} = (u_1, u_2, \ldots, u_N)^T \tag{1.1}$$

*Corresponding author
Email addresses: entidhar.malik@yahoo.com (Entidhar Malik Hadi), reham.ihsan@yahoo.com (Reham Ihssan Kamal), ali.majeed@esraa.edu.iq (Ali M. Ahmed)

Received: March 2021 *Accepted:* May 2021

Alternative H_1 is that at least part of the component of the vector under consideration is fixed in the presence of speech sounds

$$\vec{x} = \vec{u} + \vec{s}, \vec{s} = (s_1, s_2, \ldots, s_N)^T \tag{1.2}$$

It is proposed to use statistics as a decisive function in the selection of pauses [1]:

$$F_u = W_u(x).\gamma_u(x), \tag{1.3}$$

where $W_u(x)-$ characteristic that takes into account differences in values energy,
$\gamma_u(x)-$ a measure of the difference in the distribution of energy shares along the frequency axis of the compared segments.
The measure of the difference in energy values is proposed to be estimated as the ratio of the energy of the analyzed signal segment to the average energy determined on the basis of learning from a fragment corresponding to the noise in the pause:

$$W_u(x) = \frac{\| \vec{x} \|^2}{G_u}, \tag{1.4}$$

where $\| \vec{x} \|^2$ – energy (squared Euclidean norm) of the analyzed vector.
G_u- expectation of squares of Euclidean norms segments of noise in pauses.

$$G_u = M[|\vec{u}|^2] \tag{1.5}$$

In turn, the measure of the difference in the distribution of shares It is proposed to determine energies by frequency intervals in accordance with the expression, which is based on the analogue of the Pitman distance [2]:

$$\gamma_u(x) = \left(\sum_{n=0}^{N/2-1} \left((Pd_n(\vec{x}))^{1/2} - D_n \right)^2 \right)^{1/2} = \left(2 \left(1 - \sum_{n=0}^{N/2-1} D_n \left(Pd_n(\vec{x})^{1/2} \right) \right) \right)^{1/2} \tag{1.6}$$

Where $Pd_n(\vec{x})-$ the value of the fraction of energy concentrated in the$n-th$ frequency t interval:

$$Pd_n(\vec{x}) = P_n(\vec{x}) / \sum_{n=0}^{N/2-1} P_k(\vec{x}), \qquad n = 0, 1, \ldots, N/2-1, \tag{1.7}$$

D_n^2- mathematical expectation of the energy fractions of noise segments in pauses

$$D_n^2 = M[Pd_n(\vec{u})], \qquad n = 0, 1, \ldots, N/2-1 \tag{1.8}$$

$N-$ analysis interval duration.
Figures 1-3 show a fragment of the PC generated by the word "turtle" and the result of evaluating characteristics (1.4) and (1.6). In this case, the value of the mathematical expectations D_n^2 and G_u was determined based on the analysis of a noise fragment in a pause at the beginning of a signal fragment with a duration of 0.19 sec.

The function $W_u(x)$ reacts to a change in energy compared to the average, while $\gamma_u(x)$ reacts to a change in its distribution over frequency intervals. The hypothesis H_0 is rejected if the inequality

$$F_u > h_\alpha \tag{1.9}$$

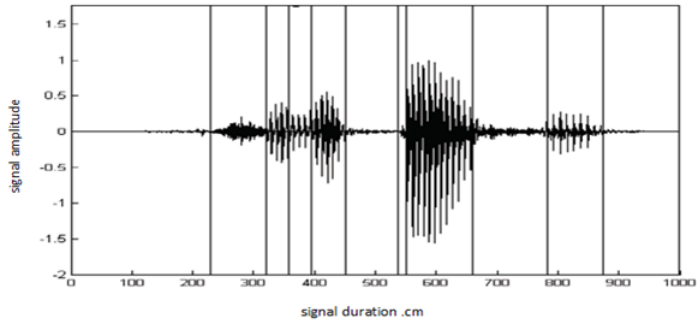

Figure 1: Fragment of the PC generated by the word "turtle" (fd = 16 kHz).

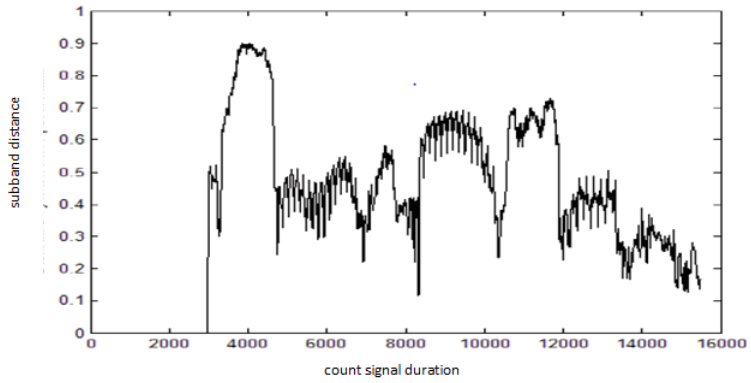

Figure 2: The result of the estimation of the subband distance $\gamma_u(x)$ of the PC fragment generated by the word "turtle" (fd = 16 kHz, N = 256)

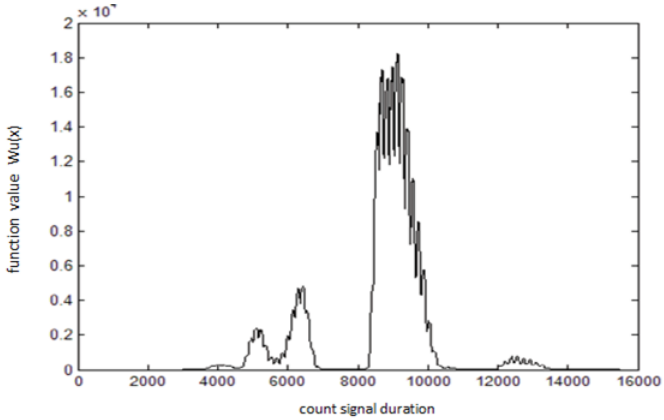

Figure 3: The result of evaluating the function$W_u(x)$ of the PC fragment generated by the word "turtle" (fd = 16 kHz, N = 256).

whre $h_\alpha > 0-$ threshold satisfying the condition

$$PR\left(F_u > \frac{h_\alpha}{H_0}\right) \leq \alpha \tag{1.10}$$

Here PR is the symbol of probability, and α is the desired level the probabilities of errors of the first kind. To assess the effectiveness of the developed algorithm, estimates of the probabilities

of errors of the first and second kind were used. The estimate of the probability of a type I error was determined based on the analysis of the signal corresponding to the noise section in the pauses (185000 samples). The probability value was defined as:

$$P_{1osch} = \frac{N_{\text{felt speech}}}{N_{iauz}}, \tag{1.11}$$

where
$N_{\text{felt speech}}$ – the number of segments erroneously assigned to RS in the presence of speech sounds.
N_{iauz} – the number of PC segments generated by noise, used called for analysis (185,000) segments.

To assess the likelihood of a type II error, speech material with previously removed sections of pauses (230,000 segments) was used. The error probability was determined using a relation of the form:

$$P_{2osch} = \frac{N_{\text{felt iauz}}}{N_{\text{speech}}}, \tag{1.12}$$

where
$N_{\text{felt iauz}}$ – the number of segments erroneously attributed to noise in pause,
N_{speech} –the number of PC segments in the presence of speech sounds, used called for analysis (230,000) segments.

Table 1: Values of the probabilities of errors of the first and second kind at different parameters (fd = 16kHz)

	$N = 128$		$N = 256$	
	P_{1osch}	P_{2osch}	P_{1osch}	P_{2osch}
without noise	0.0332	$< 10^{-4}$	0.0791	$< 10^{-4}$
k = 0.1	0.0374	0.0002	0.0861	$< 10^{-4}$
k = 0.2	0.0410	0.0015	0.0902	$< 10^{-4}$
k = 0.3	0.0462	0.0027	0.0963	0.0002
k = 0.4	0.0507	0.0092	0.1027	0.0006
k = 0.5	0.0537	0.0184	0.1081	0.0029
k = 0.6	0.0559	0.0298	0.1127	0.0057
k = 0.7	0.0573	0.0465	0.1164	0.0075
k = 0.8	0.0582	0.0621	0.1201	0.0109
k = 0.9	0.0584	0.0904	0.1235	0,0187
k = 1	0.0583	0.1161	0.1265	0.0256

2. Conclusions

Table 1 shows the results of evaluating the probabilities of errors for different values of the durations of the analysis segments and different ratios of noise / signal k.The results obtained show that the proposed method makes it possible to identify areas of pauses with a low probability of erroneous decision-making

References

[1] V. B. Sadov, *To a question of automatic control of the drive of the sucker rod pump*, Bull. South Ural State University, Ser. Comput. Tech. Cont. Radio Elec. 13(3) (2013) 46–53.

[2] A. Azzalini, *Statistical Inference: Based on TheLikelihood*, Routledge, 2017.
[3] L. Karray and Arnaud Martin, *Towards improving speech detection robustness for speech recognition in adverse conditions*, Speech Commun. 40(3) (2003) 261–276.
[4] J. Ramirez, J. C. Segura, C. Benítez, A. de la Torre and A. Rubio, *A new adaptive long-term spectral estimation voice activity detector*, Eighth Europ. Conf. Speech Commun. Tech. 2003.
[5] J. Ramírez, J. C. Segura, C. Benitez, A. de la Torre and A. Rubio, *An effective subband OSF-based VAD with noise reduction for robust speech recognition*, IEEE Transactions on Speech and Audio Proc. 13(6) (2005) 1119–1129.
[6] A. Benyassine, E. Shlomot, H.-Y. Su, D. Massaloux, C. Lamblin and J.-P. Petit, *ITU-T Recommendation G. 729 Annex B: a silence compression scheme for use with G. 729 optimized for V. 70 digital simultaneous voice and data applications*, IEEE Commun. Mag. 35(9) (1997) 64–73.
[7] Sangwan, Abhijeet, et al, *VAD techniques for real-time speech transmission on the internet*, 5th IEEE International Conference on High Speed Networks and Multimedia Communication (Cat. No. 02EX612). IEEE, 2002.
[8] F. Basbug, K. Swaminathan and S. Nandkumar, *Noise reduction and echo cancellation front-end for speech codecs*, IEEE Trans. Speech Audio Proc. 11(1) (2003) 1–13.
[9] S. Gustafsson, R. Martin and P. Vary, *A psychoacoustic approach to combined acoustic echo cancellation and noise reduction*, IEEE Trans. Speech Audio Proc. 10(5) (2002) 245–256.
[10] S. Bradley, J. Backman, S. von Hunerbein and T. Wu, *The mechanisms creating wind noise in microphones*, Audio Engin. Soc. Conven. 114. Audio Engin. Soc. 2003.
[11] S. Morgan, and R. Raspet, *Investigation of the mechanisms of low-frequency wind noise generation outdoors*, J. Acoust. Soc. Amer. 92(2) (1992) 1180–1183.
[12] J. Wang, P. Du, T. Niu and W. Yang, *A novel hybrid system based on a new proposed algorithm—multi-objective whale optimization algorithm for wind speed forecasting*, Appl. Energy, 208 (2017) 344–360.
[13] Q. Shakir kadhima, H. Kh. Obayes, M. H. Rashid, S. H. Abdul-zahrah and O. Khudhayer Obayes *Features of molecules accumulation in the triplet state at excitation of the organic compounds by rectangular pulses*, Organic Elect. 64 (2019) 202–204.