1200Hz, but the results were the similar as in scenarios with band pass filter.

Another approach consists in the mix between MFCCs and zero crossing rate (ZCR) or combination between the zero crossing rate (ZCR) and the short-term energy (called as ZRMSE) as the last coefficient. It was expected that these features would lead to high errors at low SNRs, but tables X and XI show that this fact is true for high SNRs.

The errors introduced in the case when ZCR or ZRMSE have been used have increased considerably.

V. CONCLUSION AND FURTHER WORK

In this paper, we proposed an algorithm to detect simultaneous speech from voice communications systems for *air* or *naval* traffic control, which is based on traditional Mel Frequency Cepstral Coefficient and Dynamic Time Warping. We introduced a 2-stage normalization procedure, which reduced missed and false detection rates. We also analyzed the use of other several speech features, simulated with TIMIT, Noisex-92 and a proprietary data base.

On future experiments we will mixt the MFCC with various speech features aiming for better results. We will also try different new combinations adding speech features presented in Section II.A. A new research direction for this task is represented by deep learning, which recently was used in [22] for overlapped speech detection.

REFERENCES

[1] L. Friedrich, “Method and device for the detection of simultaneous dual emission of AM signals”, pattent DE102007037105 A1, 2008.

[2] ED-137B - “Interoperability Standards for VoIP ATM Components”, European Organization for Civil Aviation Equipment, 2016.

[3] R. S. Marinescu, C. Burileanu, “Voice activity detection for best signal selection in air traffic management and control systems”, in the *Proc. 38th International Conference on Telecommunications and Signal Processing (TSP)*, Prague, Czech Republic, 2015.

[4] R. S. Marinescu, “Best Signal Selection with Automatic Delay Compensation in VoIP Environment”, PhD Thesis, University Politehnica of Bucharest, Romania, 2013.

[5] T. E. Tremain, “The government standard Linear Predictive Coding Algorithm: LPC10”, *Speech Technology*, Vol. 1, No.2, pp. 40-49, 1982.

[6] P. Mermelstein, “Distance measures for speech recognition, psychological and instrumental,” in *Pattern Recognition and Artificial Intelligence*, C. H. Chen, Ed., pp. 374–388. Academic, US, 1976.

[7] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” *Journal of Acoustical Society of America*, no. 87, pp. 1738-1752, 1990.

[8] R. Bellman, “Dynamic Programming,” *Princeton University Press*, 1957.

[9] M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.

[10] A. Ouzounov, “Robust Feature for Speech Detection”, *Cybernetics and Information Technologies*, vol.4, No.2, pp.3-14, Bulgaria, 2004.

[11] L. S. Huang and C. H. Yang, “A novel approach to robust speech endpoint detection in car environments”, in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1751-1754, Turkey, 2000.

[12] T. Kristiansson, S. Deligne, P. Olsen, “Voicing Features for Robust Speech Detection”, in *Proc. 6th Annual Conference of the International Speech Communication Association (ISCA)–INTERSPEECH*, pp. 369-372, Portugal, 2005.

[13] P. Boersma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” in *IFA Proceedings. Institute of Phonetic Sciences*, University of Amsterdam, pp. 97-110, 1993.

[14] D. Talkin, “Speech Coding and Synthesis”, *Elsevier Science B.V.*, 1995.

[15] T. Drugman and A. Alwan, “Joint robust voicing detection and pitch estimation based on residual harmonics,” in *Proc. Interspeech*, Italy, 2011.

[16] A. de Cheveign´e and H. Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917-1930, 2002.

[17] V. Ramasubramanian, A. Das, and V. Kumar, “Text-dependent speaker-recognition using one-pass dynamic program-ming”, in *Proc. ICASSP’06*, France, 2006.

[18] N. Murali Krishna, P.V. Lakshmi, Y. Srinivas, J. Sirisha Devi, “Emotion Recognition using Dynamic Time Warping Technique for Isolated Words”, *International Journal of Computer Science Issues*, Vol. 8, Issue 5, No 1, pp. 306-309, 2011.

[19] S. V. Chapaneri, “Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping”, *International Journal of Computer Applications*, vol, 40, no. 3, pp. 6-12, 2012.

[20] W. Fu, X. Yang, and Y. Wang, “Heart sound diagnosis based on DTW and MFCC”, *3rd International Congress on Image and Signal Processing*, pp. 2920-2923, 2010.

[21] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscuc, D. S. Pallett, N. L. Dahlgren, V. Zue, “TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1”, Web Download. Philadelphia: Linguistic Data Consortium, 1993.

[22] V. Andrei, H.Cucu, C. Burileanu, “Detecting overlapped speech on short timeframes using deep learning”, *INTERSPEECH 2017*, pp. 1198-1202, Sweden, 2017.