DE GRUYTER OPEN

Print ISSN: 1312-2622; Online ISSN: 2367-5357
DOI: 10.1515/itc-2016-0005

Mean-Delta Features for Telephone Speech Endpoint Detection

A. Ouzounov

Key Words: *Endpoint detection; speaker verification; group delay spectrum.*

Abstract. *In this paper, a brief summary of the author's research in the field of the contour-based telephone speech Endpoint Detection (ED) is presented. This research includes: development of new robust features for ED – the Mean-Delta feature and the Group Delay Mean-Delta feature and estimation of the effect of the analyzed ED features and two additional features in the Dynamic Time Warping fixed-text speaker verification task with short noisy telephone phrases in Bulgarian language.*

Introduction

The errors in the automatic speech and speaker recognition systems designed to operate in real-world environments are due to many reasons including the inaccurate detection of the endpoints of the analyzed speech utterance. The wrong Endpoint Detection (ED) increases the cases when the system processes data different from the actual speech utterance. These errors are crucial especially for recognition systems, which use short phrases with length of few seconds.

The ED algorithm consists of two main processing steps – feature extraction and decision step. In the first processing step, the features based on signal energy [3,6], spectral entropy [4,5], group delay functions [16], wavelets [19], etc., are extracted. In the second step, using the properties of the estimated features, the start and the end points of the utterance are estimated. This is accomplished by using a state automaton [6] or some type of classification scheme, e.g., classification and regression tree [18], Hidden Markov Models (HMM) [17], support vector machines [14], etc.

In the paper, a brief summary of the author's research in the field of the contour-based telephone speech ED is presented. This research work includes:

- Development of new robust features for ED: the FFT magnitude spectrum-based Mean-Delta feature [11] and the Group Delay Mean-Delta feature [12].
- Estimation of the effect of the analyzed ED features and two additional features - the modified Teager energy [4] and the energy-entropy feature [5] in the Dynamic Time Warping (DTW) fixed-text speaker verification task with short noisy telephone phrases in Bulgarian language [11,12].
- Development of thresholds setting and state automaton algorithms necessary for the ED of a single word or short utterance [11].

Further, the finite state machine based decision logic applied to the current endpoint detection is described here in detail.

The Mean-Delta (MD) Feature

The MD feature is proposed by the author in [10] and is defined as the mean absolute value of the Delta Spectral AutoCorrelation Function (DSACF) of the speech spectrum. For a particular frame, the DSACF was computed utilizing only the frame's spectral autocorrelation lags and it is obtained in a way similar to the delta cepstrum evaluation – an orthogonal polynomial fit of the first-order derivative (in correlation domain). For the *n*th frame, the DSACF $\Delta R_P(n,l)$ is

$$(1)\quad \Delta R_P(n,l) = \frac{\sum_{q=-Q}^{Q} qR_P(n,l+q)}{\sum_{q=-Q}^{Q} q^2},$$

where *l=0,...,L* is the number of correlation lags; *n=0,...,N-1*, *N* is the number of frames; *Q* is the delta window and $R_P(.)$ is the biased Spectral AutoCorrelation Function (SACF) defined with the power [10] or with the magnitude spectrum [11]. For n^{th} frame the MD feature $m_d(n)$ is computed as follows:

$$(2)\quad m_d(n) = \left[\sum_{l=0}^{L} \left|\Delta R_P^S(n,l)\right|\right]^{0.5},$$

where $\Delta R_P^S(n,l)$ is the contour smoothed DSACF for lag *l*. This smoothing is obtained by the Long-Term Spectral Envelope (LTSE) algorithm [13], applied on the DSACF. The block diagram of the algorithm for the magnitude spectrum-based MD feature extraction is shown in *figure 1*.

The Group Delay MD Feature

The Group Delay Mean Delta (GDMD) feature is proposed by the author in [12]. This feature utilized the Mean Delta approach proposed in [10] but the spectral autocorrelation function is defined based on the Modified Group Delay Spectrum (MGDS), instead on the magnitude spectrum. The aim of this is to obtain peak-enhanced delta spectral autocorrelation function and thereafter more effective Mean Delta feature.

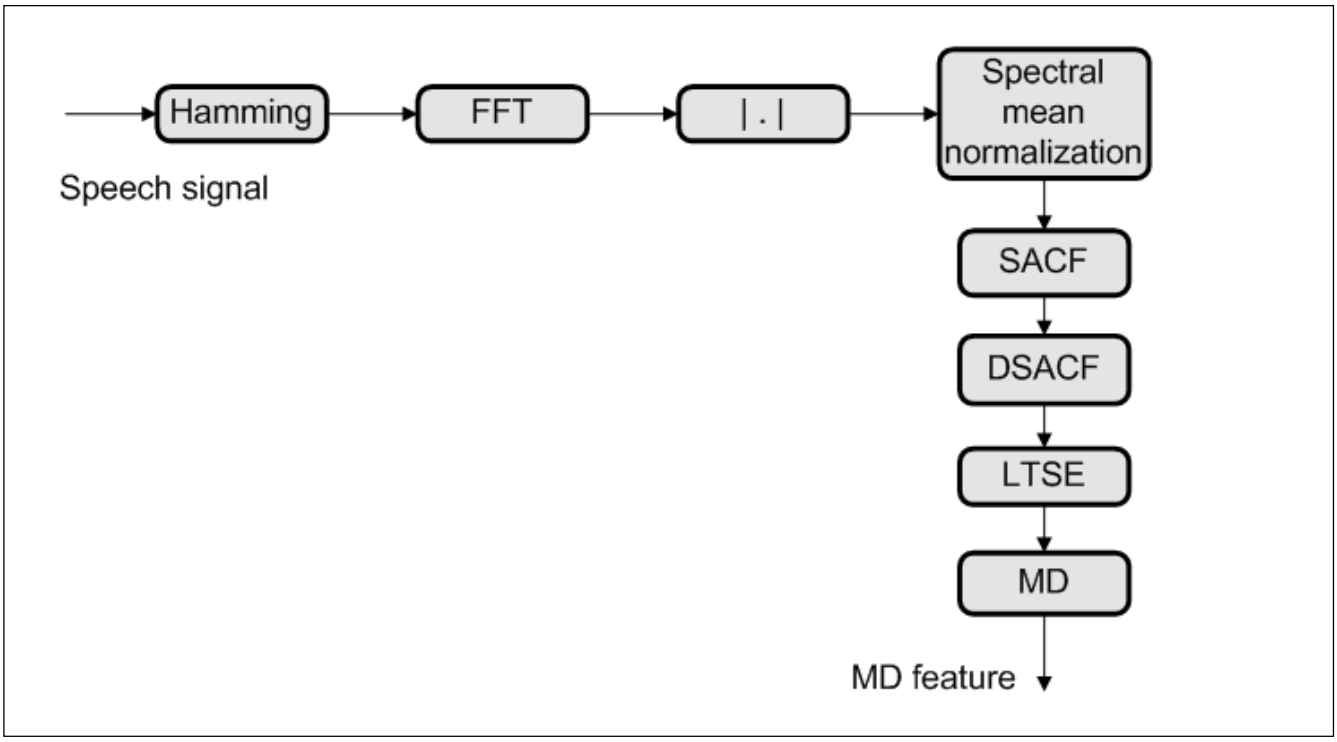

Figure 1. Block diagram of the algorithm for the magnitude spectrum-based MD feature extraction

The MGDS $\tau_m(k)$ is proposed in [7] and is defined as

$$\tau_m(k) = sign.\left|\frac{X_R(k)Y_R(k)+Y_I(k)X_I(k)}{S(k)^{2\gamma}}\right|^{\alpha}, \quad (3)$$

where *sign* is given by the sign of the term in the absolute value brackets; x(n) is the given speech frame; X(.) and Y(.) are the Fourier transforms of the sequences x(n) and nx(n); $k = 0, \ldots, K/2$; ; K is the FFT size; S(.) is the cepstrally smoothed spectrum of |X(.)| using low-order cepstral lifter l_w. α, γ and l_w are adjusted according to the particular requirements [7]. The GDMD feature is computed in the same way as the MD one, but instead the FFT magnitude spectrum the MGDS is used [12].

The Modified Teager Energy (MTE) Feature

The MTE feature $E_t(n)$for n^{th} frame is [4]

$$E_t(n) = \left[\sum_{k=0}^{K/2}(k\Delta f)^2\left|X(n,k)\right|^2\right]^{0.5}, \quad (4)$$

where Δf is the frequency resolution, $|X(n,k)|^2$ is the FFT power spectrum and K is the FFT size;

The Energy Entropy (EE) Feature

The EE feature is obtained by combination of the energy and the spectral entropy and for n^{th} frame is defined as [5]

$$EE(n) = \sqrt{(1+\left|E(n)H(n)\right|)}, \quad (5)$$

where $E(n)$ is the frame energy and $H(n)$ is the frame spectral entropy. The probability density function $P(n, k)$ for the frequency component k is

$$P(n,k) = \frac{\left|X(n,k)\right|^2}{\sum_{k=0}^{K/2}\left|X(n,k)\right|^2}, \quad (6)$$

and $H(n)$ is

$$H(n) = -\sum_{k=0}^{K/2}P(n,k).\log(P(n,k)). \quad (7)$$

Endpoints Detection Algorithm

The proposed Endpoint Detection (ED) algorithm [11] is intended for location of beginning and ending frames of a word or single utterance with a short length (few seconds). The block diagram of the ED algorithm is shown in *figure 2*.

In this study the fixed thresholds approach has been used. The aim of the fixed threshold is to separate the noise frames from the noise and speech frames based only on the value of selected parameter. The method proposed in [3] is based on the observation that the histograms of the log energy of noisy speech have a clean bimodal distribution corresponding to "noise only" and "noise + speech" parts of signal. In this case the distribution can be approximated with two Gaussian densities that allow deriving statistically

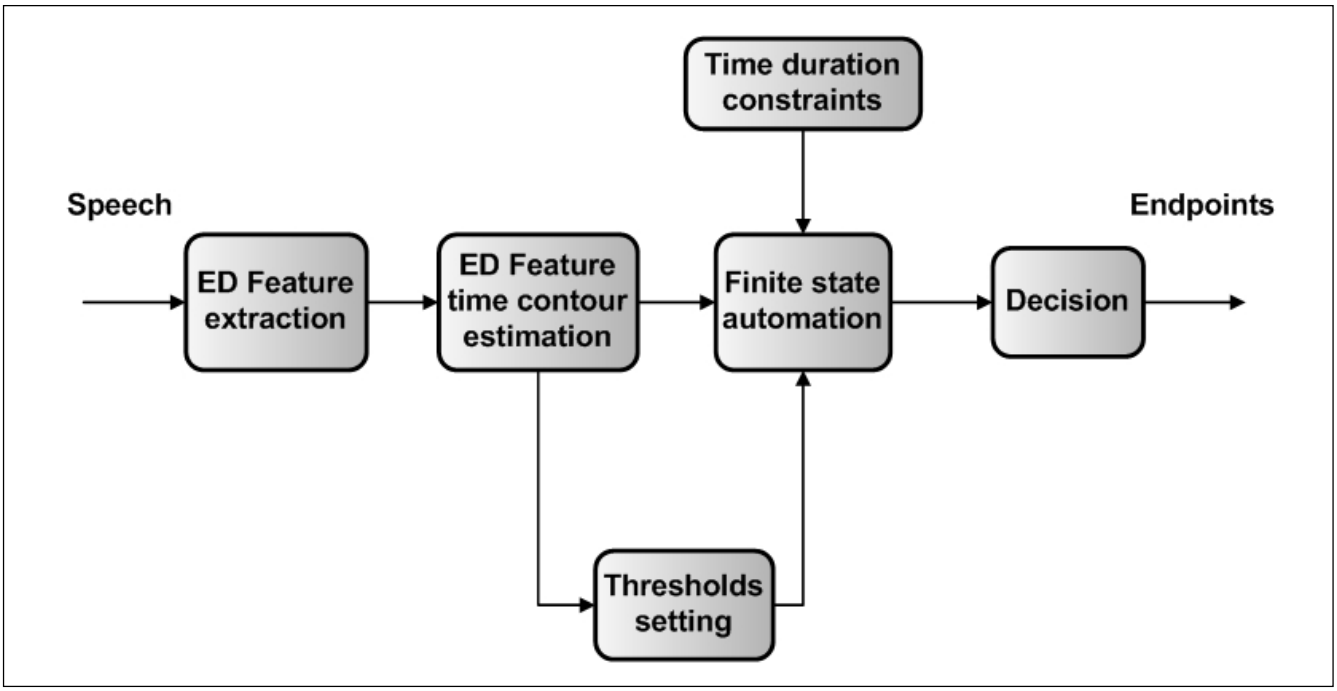

Figure 2. Block diagram of the ED algorithm

optimal threshold applying the EM (Expectation-Maximization) algorithm [3]. It is supposed that speech dominates the noise and this modeling is suitable only for cases with significant positive Signal-to-Noise Ratio (SNR) and stationary background noise [3]. Since in the real-world environments (e.g. telephone speech) it is difficult to meet similar conditions, this algorithm is not used in the study.

The simpler algorithm with two fixed thresholds (T_{low} and T_{high}) is proposed by the author in [11]. In the text below is described its improved version used in [12]. The preliminary experiments had shown its reliable work in moderate noise levels. In this algorithm is utilized a base threshold which is equal to the average value computed over entire utterance contour. Through the use of the base threshold two additional average values m_{down} and m_{up} are estimated. The first value is the average of the contour values that are less than the base threshold and the second one is the average of the values that are equal to or greater than the base threshold. The low threshold T_{low} is defined as a sum of m_{down} and a part of the difference between m_{up} and m_{down}. The high threshold is defined as $T_{high} = \beta T_{low}$. The coefficients α, β and γ are experimentally determined and their typical values are $\alpha = 0.03$, $\beta = 1.5$ and $\gamma = 0.05$. The flowchart of the algorithm is shown in *figure 3*.

The proposed ED algorithm is based on eight-state automaton and it will be described in detail. The eight states are: *INIT, SCAN_DATA, SCAN_START, MAYBE_IN, SCAN_END, MAYBE_OUT, END_FOUND* and *END*. The transition from one state to another is controlled by the rules based on the two thresholds scheme and some duration constraints. These constraints are included in order to filter (to some extent) prolonged low-level and short high-level non-speech events before and after the speech utterance. Also, it is supposed that the utterance starts and ends within the audio file.

Main purposes of the states are:

- o *SCAN_DATA* – Search for beginning point candidate;
- o *SCAN_START* – Scan data between two thresholds;
- o *MAYBE_IN* – Estimate the beginning point;
- o *SCAN_END* – Search for ending point candidate;
- o *MAYBE_OUT* – Estimate the ending point;
- o *END_FOUND* – Check the utterance length.

The finite state machine based decision logic applied for ED is shown in *figure 4*. The rules of the state transition in the state machine are presented in *figure 5*. If an error occurs, the ED algorithm stops and the particular file will be ignored in further processing. This can occur in two cases. First, when the utterance starts or/and ends outside the audio file (ERR_TOOSHORT, ERR_TOOLONG), and second, when the SNR is very low (ERR_NOSPEECH, ERR_LOWSPEECH). For the sake of clarity, the errors are not shown in *figure 4*.

In *figure 5* the parameters T_{SCAN_DATA}, T_{SCAN_START}, T_{MAYBE_IN}, T_{SCAN_END} and T_{MAYBE_OUT} are the corresponded state timers. Each one of the time constants *MaxWaitTime, MaxQuietTime, BegTime, MaxEndTime, UpTime, MiddleTime, MinLengthTime* determines the length of the interval after which the specified state transition will occur. Their values are set according to the particular requirements.

For illustration in *figure 6* are shown the contours of the described above features for a street noise example (with sound of a car alarm) with SNR=0 dB selected from the NOIZEUS corpus [20,21]. The example has clean speech reference and corresponded noisy version (time-aligned). In *figure 6 (c), (d), (e)* and *(f)* are shown the features' con-

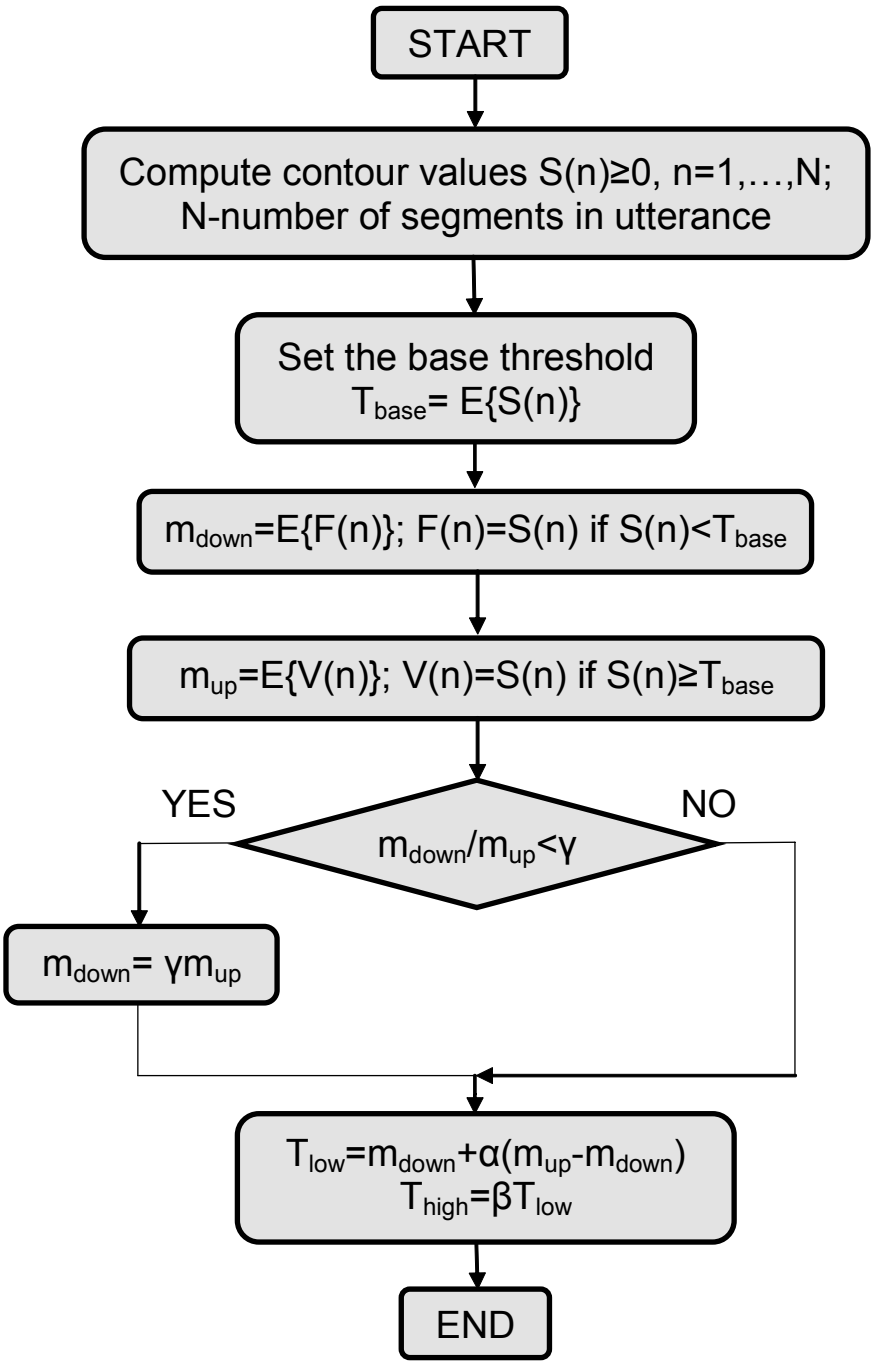

Figure 3. Flowchart of the two thresholds setting algorithm

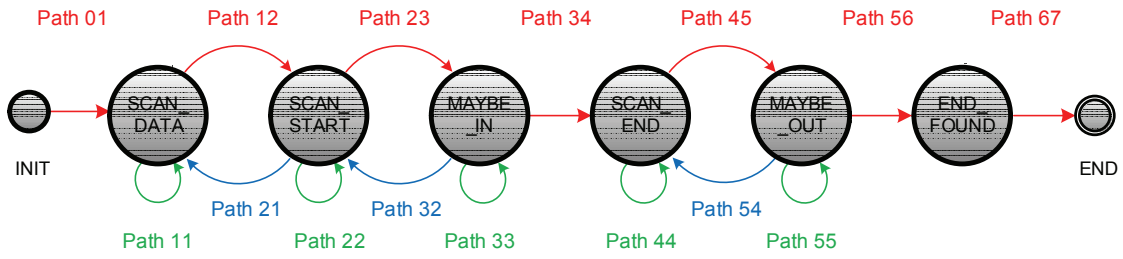

Figure 4. Finite state machine based decision logic diagram

Paths	State transition	Rules of state transition	Errors
Path 01	INIT→ SCAN_DATA	Go to SCAN_DATA after all parameters being set to default values.	
Path 11	SCAN_DATA → SCAN_DATA	Stay in SCAN_DATA, if $(E(n) \le T_{low})$ & $(T_{SCAN\_DATA} < MaxWaitTime)$.	If $(E(n) \le T_{low})$ & $(T_{SCAN\_DATA} \ge MaxWaitTime)$- ERR_NOSPEECH.
Path 12	SCAN_DATA→ SCAN_START	Go to SCAN_START, if $(E(n) > T_{low})$ - ***n*** is marked as beginning point candidate.	
Path 21	SCAN_START→ SCAN_DATA	Go back to SCAN_DATA, if $(E(n) \le T_{low})$.	
Path 22	SCAN_START→ SCAN_START	Stay in SCAN_START if $(E(n) > T_{low})$ & $(E(n) \le T_{high})$ & $(T_{SCAN\_START} < MaxQuietTime)$.	If $E(n) > T_{low})$ & $(E(n) \le T_{high})$ & $(T_{SCAN\_START} \ge MaxQuietTime)$ -ERR_LOWSPEECH.
Path 23	SCAN_START→ MAYBE_IN	Go to MAYBE_IN, if $(E(n) > T_{high})$.	
Path 32	MAYBE_IN → SCAN_START	Go back to SCAN_START, if $(E(n) \le T_{high})$ & $(T_{MAYBE\_IN} < BegTime)$.	
Path 33	MAYBE_IN → MAYBE_IN	Stay in MAYBE_IN, if $(E(n) > T_{high})$ & $(T_{MAYBE\_IN} < BegTime)$.	
Path 34	MAYBE_IN → SCAN_END	Go to SCAN_END, if $(E(n) > T_{high})$ & $(T_{MAYBE\_IN} \ge BegTime)$ Estimate the beginning point - *BPoint*.	
Path 44	SCAN_END → SCAN_END	Stay in SCAN_END, if $(E(n) > T_{low})$ & $(T_{SCAN\_END} < MaxEndTime)$.	If $(E(n) > T_{low})$ & $(T_{SCAN\_END} \ge MaxEndTime)$ - ERR_TOOLONG.
Path 45	SCAN_END → MAYBE_OUT	Go to MAYBE_OUT, if $(E(n) \le T_{low})$ - ***n*** is marked as ending point candidate.	
Path 54	MAYBE_OUT→ SCAN_END	Go back to SCAN_END, if $((E(n) > T_{high})$ & $(T_{SCAN\_END} > UpTime))$ OR $((E(n) \le T_{high})$ & $(E(n) > T_{low})$ & $(T_{SCAN\_END} > MiddleTime))$.	
Path 55	MAYBE_OUT → MAYBE_OUT	Stay in MAYBE_OUT, if $(E(n) \le T_{low})$ & $(T_{MAYBE\_OUT} \le MaxEndTime)$.	
Path 56	MAYBE_OUT→ END_FOUND	Go to END_FOUND, if $(E(n) \le T_{low})$ & $(T_{MAYBE\_OUT} > MaxEndTime)$ Estimate the ending point - *EPoint* and the utterance length *ULength =EPoint-BPoint.*	
Path 67	END_FOUND→ END	Go to END, if $(ULength \ge MinLengthTime)$.	If $(ULength < MinLengthTime)$ - ERR_TOOSHORT.

Figure 5. The rules of the state transition

tours of the noisy example in *figure 6 (a).* Also are shown the thresholds and endpoints (vertical red lines) estimated according to the detection algorithm described above. It can be seen in *figure 6 (c)* that for this noisy example the MTE feature is not suitable for contour-based endpoint detection.

Speech Data

The speech data used in the experiments are selected from the BG-SRDat corpus [9]. This corpus is in Bulgarian language and it is recorded over noisy telephone channels and intended for speaker recognition. The data are sampled with frequency of 8 kHz at 16 bits, PCM format, and mono mode. The length of the selected utterance is about 2 seconds and the length of the single record (file) is about 2.5-3 seconds.

It is worth to make some clarifications about the used phrase in Bulgarian language. It starts with voiced fricative *'z'* and ends with unvoiced fricative*'s'*. The phrase is: "*Zdravei Manolov. Kak se chuvstvash dnes?"*. Its English meaning is "*Hello Manolov! How are you today?"*. The pronunciation (roughly) is"*[zdra`vei:] [ma`nolov]! [kak] [se] [`tʃuvstvaʃ] [dnes]?"*[9]. In addition, the manual labelling of the endpoints of all speech data is done in order to have reference endpoints for comparative purposes.

Speaker Verification Performance

Speaker modeling in the password-based speaker verification system can typically be done in two ways. The former is in the signal domain using DTW algorithm, and the latter is with stochastic model of the speaker using HMM approach. In the paper the proposed endpoint detector is examined as a part of the fixed-text DTW-based speaker verification system. Its evaluation in the HMM framework for short phrases will be described in the forthcoming paper of the author.

The speech data used in the study include 262 records of a phrase collected from 12 male speakers. As the speech corpus is not large enough we cannot use two separate data set in training mode – one for reference template creation (training set) and another for thresholds setting (validation set). Therefore, in the study the training set is used directly as a validation set. The different numbers of records per speaker (from 16 up to 34) and requirements to use

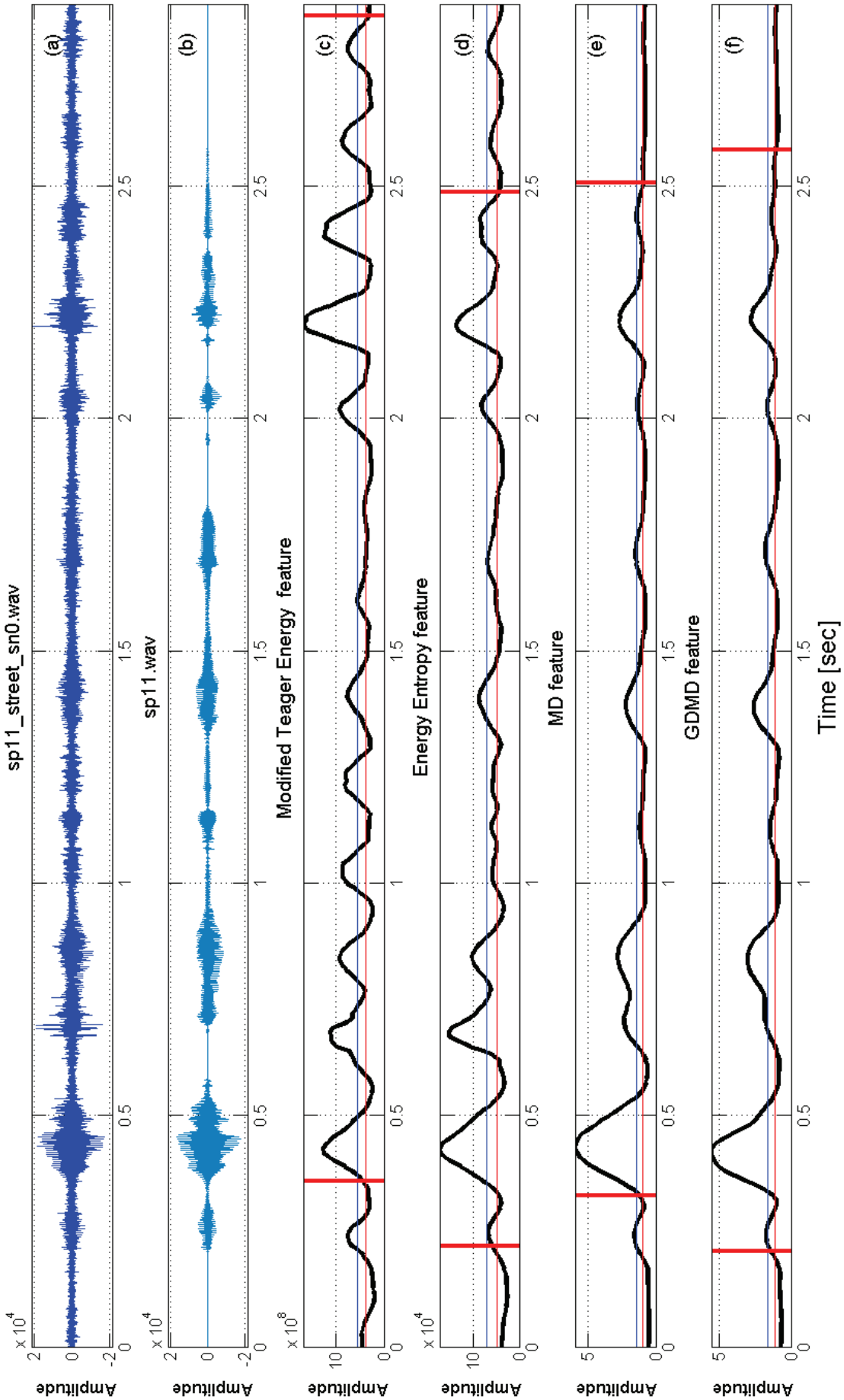

Figure 6. Examples from the NOIZEUS corpus: (a) noisy example; (b) the clean version; (c) modified Teager energy; (d) energy entropy feature; (e) MD feature; (f) GDMD feature

Speaker verification results

№	Features	FRR[%]	FAR[%]	HTER[%]	95% CI
1	**Manual**	6.90	4.98	5.94	±0.0096
2	**MTE**	11.83	10.47	11.15	±0.0123
3	**EE**	14.08	12.48	13.28	±0.0133
4	**MD**	10.56	8.06	9.31	±0.0116
5	**GDMD**	8.30	7.31	7.80	±0.0105

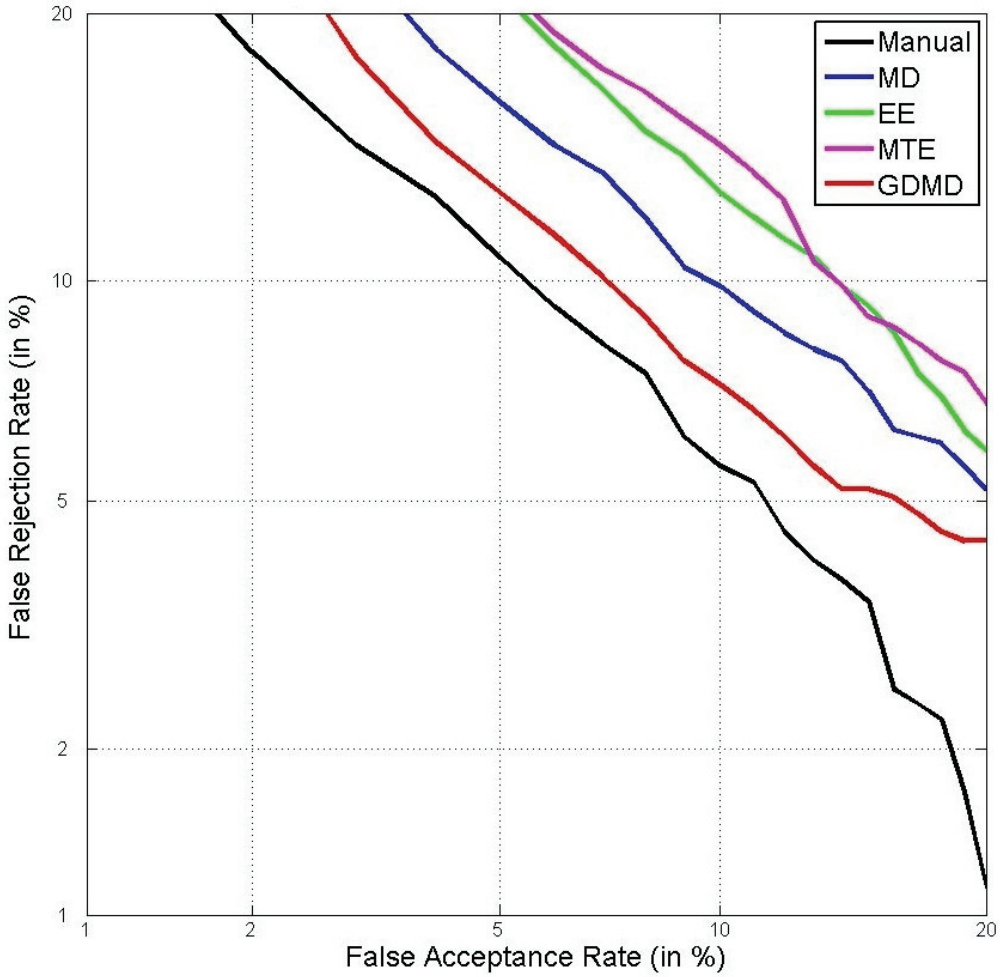

Figure 7. DET curves for different ED features

equal number of records for speaker's reference creation impose the following training procedure [12]. For reference creation are randomly selected 10 records per speaker. The rest of speaker's data are used for testing. This procedure is repeated 5 times. In the verification mode there are 142 client accesses or false rejection tests and 1562 impostor accesses or false acceptance tests. After 5 runs the total tests are: for false rejection – 710 and for false acceptance – 7810. In the pre-processing step the MEL cepstrum with 14 coefficients is used.

In the study, the normalize-wrap DTW algorithm with the root power sum – cepstral distance is applied [8]. In this algorithm are used the constrained endpoints conditions [8]. The speaker's reference is obtained by averaging (after dynamic time warping alignment) of his training utterances. The individual speakers' verification thresholds are estimated by using of the cohort normalization method [2].

The verification results are presented as rate ratios – False Rejection Rate (FRR), False Acceptance Rate (FAR) and the Half Total Error Rate (HTER) [1]. Also the 95% Confidence Interval (CI) for the HTER is shown computed according to [1]. In *the table* are shown the speaker verification results in rates and confidence interval for the HTERs. These rates are obtained for each feature and also for the manual end pointing. As seen in the table the GDMD feature performs the best among the features set.

The average DET curves [15] are plotted in *figure 7* to show the verification performance for each ED feature. It is clearly seen that the GDMD feature curve is closer to the reference curve (manual ED) than the other three ones.

Conclusions

A brief summary of the author's research in the field of the contour-based telephone speech ED was presented in the paper. Further, was described in detail the finite state machine based decision logic applied for the current endpoint detection.

Based on the research work the following conclusions are made:

- The GDMD feature demonstrates the best performance in the endpoint detection tests based on the verification rate. This is due to the minimal number of the serious endpoint detection errors obtained for this feature.
- The GDMD feature has two drawbacks: a need to adjust a few parameters in the MGDS estimation and increased number of computation (in comparison with the MD feature).

Future work in this area will be focused on three main objectives – the development of more efficient version for the GDMD feature with fewer adjustable parameters, the improvement of the endpoint detection accuracy especially for weak phonemes and the examination of the developed endpoint detector in the HMM framework for short phrases.

References

1. Bengio, S., J. Mariethoz. A Statistical Significance Test for Person Authentication. ODYSSEY – the Speaker and Language Recognition Workshop, 2004, 237-244.
2. Burileanu, C., D. Moraru, L. Bojan, M. Puchiu, A. Stan. On Performance Improvement of a Speaker Verification System Using Vector Quantization, Cohorts and Hybrid Cohort-World Models. – *International Journal of Speech Technology*, 2002, No. 5, 247–257.
3. Gerven, S., F. Xie. A Comparative Study of Speech Detection Methods. Eurospeech, 1997, 1095-1098.
4. Gu, L., S. Zahorian. A New Robust Algorithm for Isolated Word Endpoint Detection. – *IEEE ICASSP*, IV, 2002, 4161-4164.
5. Huang, L., C. Yang. A Novel Approach to Robust Speech Endpoint Detection in Car Environment. – *IEEE ICASSP*, 2000, 1751-1754.
6. Li, Q., J. Zheng, A. Tsai, Q. Zhou. Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition. – *IEEE Transaction on SAP*, 10, 2002, No. 3, 146-157.
7. Murthy, H., V. Gadde. The Modified Group Delay Function and its Application to Phoneme Recognition. – *IEEE ICASSP*, 1, 2003, 68-71.
8. Myers, C., L. Rabiner, A. Rosenberg. Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition. – *IEEE Transactions on ASSP*, 28, 1980, No. 6, 623-635.
9. Ouzounov, A. BG-SRDat: A Corpus in Bulgarian Language for Speaker Recognition over Telephone Channels. – *Cybernetics and Information Technologies,* 3, 2003, No. 2, 101-108.
10. Ouzounov, A. A Robust Feature for Speech Detection. – *Cybernetics and Information Technologies,* 4, 2004, No. 2, 3-14.
11. Ouzounov, A. Telephone Speech Endpoint Detection Using Mean-Delta Feature. – *Cybernetics and Information Technologies,* 14, 2014, No. 2, 127-139.
12. Ouzounov, A. Noisy Speech Endpoint Detection Using Robust Feature. Springer International Publishing Switzerland 2014, V. Cantoni et al. (Eds.), BIOMET 2014, LNCS 8897, 2014, 105-117.
13. Ramirez, J., et al. Efficient Voice Activity Detection Algorithms Using Long-Term Speech Information. – *Speech Communication*, 42, 2004, No. 3-4, 271-287.
14. Ramirez, J., et al. SVM-based Speech Endpoint Detection Using Contextual Speech Features. – *Electronics Letters,* 42, 2006, No. 7, 426-428.
15. Martin, A., et al. The DET Curve in Assessment of Detection Task Performance. Eurospeech, 1997, 1895-1898.
16. Krishnan, S., et al. Robust Voice Activity Detection Using Group Delay Functions. IEEE International Conference on Industrial Technology, 2006, 2603-2607.
17. Zhang, Z., S. Furui. Noisy Speech Recognition Based on Robust End-point Detection and Model Adaptation. – *IEEE ICASSP,* 1, 2005, 441-444.
18. Shin, W., B. Lee, Y. Lee, J. Lee. Speech/non-speech Classification Using Multiple Features for Robust Endpoint Detection. – *IEEE ICASSP,* 2000, 1399-1402.
19. Seok, J., K. Bae. A Novel Endpoint Detection Using Discrete Wavelet Transform. – *IEICE Transaction on Inf. & Syst.,* E82-D, 1999, No. 11, 1489-1491.
20. Hu, Y. and P. Loizou. Subjective Evaluation and Comparison of Speech Enhancement Algorithms. *Speech Communication*, 49, 2007, 588-601.
21. http://ecs.utdallas.edu/loizou/speech/noizeus/.

Manuscript received on 02.02.2016

Atanas Ouzounov *is Research Associate in the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences.*
His main research interests are in the field of speech processing and speaker recognition.

Contacts:
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad G. Bonchev St., bl. 2, 1113 Sofia
e-mail: atanas@iinf.bas.bg

INFORMATION FOR AUTHORS

The SAI Journal "**Information Technologies and Control**" publishes high-quality papers on the theory, design and application of Control Systems and Information Systems, and their components. Original papers that has not been published or accepted for publication by other journals will only be considered. Full versions of papers which have been included in proceedings of conferences and symposia could also be submitted.

Submitted manuscripts must be typewritten in English. They should not exceed 20 double-spaced taped A4 pages including figures in at least 12-point format.

Phone ++ 3592 9876169, fax ++ 3592 9876169, e-mail: sai.bg.office@gmail.com

MS Word and PDF files are **the preferred format.** Papers (MS Word and PDF files) may be submitted for publication by sending an electronic mail to the editor at sai.bg.office@gmail.com. For PDF files the authors should provide all the fonts needed to print the paper.

Style for the Paper. The first page must include the title, name(s), affiliation(s), mailing address(es), telephone number(s), e-mail address(es) and facsimile number(s) of the author(s); an Abstract not exceeding 150 words; Keywords immediately following abstract, and not exceeding five; preferred address for correspondence; footnotes with acknowledgment for support (if desired).

The Introduction should describe the purpose and the contribution of the paper. **A Conclusion** section should summarize the main results, advantages, limitations and possible applications.

A descriptive, intuitive explanation is preferred to extended formal development (theorems, lemmas, etc) when technology components and applications are presented. Detailed mathematical derivations should be put in appendixes.

References should appear as a separate bibliography at the end of the paper, numbered in square brackets, e.g., [10]. (See examples of form of appearance of references to papers in periodicals, to proceedings and to books in this issue). References to works published in languages other than English should be translated into English with the original languages given in brackets after the translated text, e.g. (in Russian).

Illustrations. Electronic version in EPS (Encapsulated Postscript) is preferred. Photos must be glossy prints, of good contrast and reasonable size (no larger than 22x28 cm). Hardcopy of illustrations should be sharp and of good contrast.

Accepting a manuscript. Authors of accepted manuscripts are required to provide the final electronic version of the paper in MS Word format preferably together with PDF and Encapsulated Postscript files for figures, along with biographies (not exceeding 100 words) and photographs of all authors. The electronic version may be submitted by e-mailing it to: sai.bg.office@gmail.com. The e-mail message should include the corresponding author's name and the type of file.

Proofreading. Page proofs for papers will be send to the corresponding authors for proofreading, usually by e-mail. The author should send the revised paper back within 14 days.

Charges. After a manuscript has been accepted for publication the author(s) will be approached with a request to pay a charge of 70 euros per paper to cover partially the cost of publication. The payment of this charge is a necessary prerequisite for publication.

Contacts. In case of any questions about the described procedures, please send e-mail to: sai.bg.office@gmail.com. Phone ++ 3592 9876169, fax ++ 3592 9876169.