AUDIO SIGNAL CLASSIFICATION IN REVERBERANT ENVIRONMENTS BASED ON FUZZY-CLUSTERED AD-HOC MICROPHONE ARRAYS

Sebastian Gergen, Anil Nagathil, Rainer Martin

Institute of Communication Acoustics, Ruhr-University Bochum, Germany
email: {sebastian.gergen, anil.nagathil, rainer.martin}@rub.de

ABSTRACT

Audio signal classification suffers from the mismatch of environmental conditions when training data is based on clean and anechoic signals and test data is distorted by reverberation and signals from other sources. In this contribution we analyze the classification performance for such a scenario with two concurrently active sources in a simulated reverberant environment. To obtain robust classification results, we exploit the spatial distribution of ad-hoc microphone arrays to capture the signals and extract cepstral features. Based on these features only, we use unsupervised fuzzy clustering to estimate clusters of microphones which are dominated by one of the sources. Finally, signal classification based on clean and anechoic training data is performed for each of the cluster. The probability of cluster membership for each microphone is provided by the fuzzy clustering algorithm and is used to compute a weighted average of the feature vectors. It is shown that the proposed method exceeds the performance of classification based on single microphones.

Index Terms— Ad-hoc Microphone Array, Clustering, Classification, Cepstral Features, LP-CMRARE, MFCC

1. INTRODUCTION

Feature-based classification of audio signals into predefined categories, e.g. speech, music or noise and into subcategories, e.g. gender of a speaker or the music genre, is an important component in audio signal processing algorithms for applications like hearing aids, mobile phones or surveillance scenarios [1]. Often, investigations in this field are based on anechoic and clean audio signals. In a more realistic scenario an audio signal emitted by one source often is contaminated by room dependent reverberation and additive signals from other sources. However, the effect of these signal degradations on the classification performance has been rarely studied. For instance in [2] a system for pitch based speech/non-speech discrimination and a further classification of the non-speech signals is presented and evaluated for an anechoic and a reverberant single source scenario. In our contribution we analyze the classification performance of audio signals in a reverberant scenario and we consider a spatial ad-hoc distribution of multiple sources and multiple receivers. As a consequence the signal-to-distortion ratio (SDR) of the captured signals varies according to the position of the sources and the microphones. Ad-hoc microphone arrays increasingly attract the attention of researchers as they aim at combining mobile devices like mobile phones, tablet computers and laptops, all of which provide audio capturing devices, the ability of audio signal processing and integrated wireless connectivity. Whereas in some investigations the goal is to accurately estimate the position of such devices for example using voice activity detection and coherence models [3], energy decay information [4], calibration signals [5] and even available compass information [6], we aim at building sub-arrays, which cluster the microphones into groups based on their similarity in the feature domain. Then, a supervised signal classification based on anechoic and clean training data determines the type of audio source which dominates each of the sub-arrays. The obtained information can be used to specifically pick one of the available signals, for example as target signal- or as a noise reference for the previously mentioned applications.

The remainder of this paper is organized as follows. In Section 2 we introduce the basic idea of our algorithm and all necessary components. The experiments and settings are explained in Section 3. We present and discuss the results in Section 4 and Section 5, before we finally conclude the paper in Section 6.

2. SYSTEM CONCEPT AND FEATURES

In a scenario of n audio sources and multiple receivers in a room, we aim at clustering microphones into n clusters, each being dominated by one of the n sources. For this purpose, our first goal is to extract features from the audio signals which allow for an unsupervised estimation of clusters. As the feature extraction is executed on each capturing device itself the amount of data transmitted to other devices or to one central device can be relatively small. The information provided by the audio features is then used for unsupervised cluster analysis, which might be done in one central device. Finally, the extracted features of the receivers which are assigned to each of the n source related clusters are used in a supervised classification step (Fig. 1). The training data of the classifier is composed of clean and anechoic signals to avoid room-dependent training which would

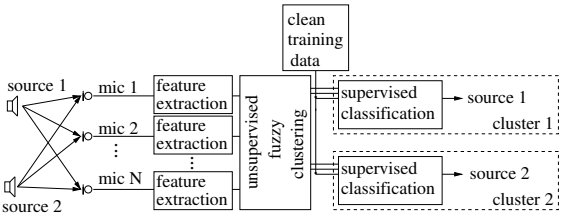

Fig. 1. Algorithm architecture for the example of two source signals. The signals are captured by several microphones and used for feature extraction. After an unsupervised clustering step, a supervised classification is performed in each cluster.

The work of S. Gergen has been supported by the Ministry of Economic Affairs and Energy of the State of North Rhine-Westphalia (Grant IV.5-43-02/2-005-WFBO-009). The work of A. Nagathil is funded by the German Research Foundation (DFG), Sonderforschungsbereich 823, Teilprojekt B3.

be too restrictive for practical applications. To make the classification results more robust, weighted means of the feature vectors of each of the devices in a cluster are evaluated.

2.1. Feature Extraction

Rather than performing classification on the captured audio signals directly, the data is typically transformed to a reduced parametric representation. In this contribution we consider two cepstrum based feature sets: the *Legendre Polynomial-based Cepstral Modulation RAtio REgression* (LP-CMRARE) and the *Modulation Mel-Frequency Cepstral Coefficients* (Mod-MFCCs) features. These features have proven to give very good results in the context of (anechoic) speech/music/noise classification tasks [7] and constitute a very compact representation of the signals.

For both feature extraction methods, first a captured audio signal $x(t)$ is sampled with the sampling rate f_s. The digital representation $x(l)$, where l is the discrete time index, is segmented into B possibly overlapping frames of length N using a window function $\mathcal{W}(n)$, e.g. the Hann window, where $n = 1, 2, \ldots, N$, and a frame shift P. The discrete Fourier transform (DFT) of the weighted frame results in the short-time Fourier transform (STFT) $X(k, b) = \mathrm{STFT}\{x(l)\}$ where b and $k = 0, 1, \ldots, N-1$ denote the frame index and the frequency bin, respectively.

2.1.1. LP-CMRARE

To obtain the LP-CMRARE features, the spectrum $X(k, b)$ is transformed into the cepstral domain $X_\mathrm{c}(\ell, b)$, where $\ell = 0, 1, \ldots, N-1$ is the index of the cepstral coefficient. Since the cepstrum is symmetric with respect to $\ell = \frac{N}{2} + 1$, $X_\mathrm{c}(\ell, b)$ is only considered for $\ell = 0, 1, \ldots, \frac{N}{2} + 1$ in the following. To analyze the spectro-temporal evolution of the cepstrum a sliding window DFT is used to compute the time-varying modulation spectrum of the cepstrum,

$$\hat{X}_\mathrm{c}(\nu, \ell, c) = \sum_{m=0}^{M-1} X_\mathrm{c}(\ell, cQ + m) e^{-j\frac{2\pi\nu m}{M}}, \qquad (1)$$

where, starting at sub-frame index $b = cQ$, the sliding window considers M consecutive sub-frames. The modulation frequency bin index is specified by $\nu = 0, 1, \ldots, M-1$ and c and Q depict the modulation window index and shift, respectively [7]. The magnitude of the modulation spectrum is averaged over all modulation analysis windows C_T,

$$\tilde{X}_\mathrm{c}(\nu, \ell) = \frac{1}{C_T} \sum_{c=0}^{C_T-1} |\hat{X}_\mathrm{c}(\nu, \ell, c)|, \qquad (2)$$

and approximately represented as cepstral modulation ratios (CMR), where the average of the modulation frequency bands $\nu_1 \leq \nu \leq \nu_2$ is normalized on the zeroth modulation frequency band (3),

$$r_{\nu_1|\nu_2}(\ell) = \frac{\sum_{\nu=\nu_1}^{\nu_2} \tilde{X}_\mathrm{c}(\nu, \ell)}{(\nu_2 - \nu_1 + 1)\tilde{X}_\mathrm{c}(0, \ell)}. \qquad (3)$$

This representation can be parametrized efficiently by fitting Legendre polynomials of order p, which finally gives $p + 1$ LP-CMRARE parameters [8].

2.1.2. Mod-MFCC

To compute Mod-MFCC coefficients, the magnitude squared spectral representation $|X(k, b)|^2$ is mapped onto the mel scale using overlapping triangular windows [9]. The resulting mapped spectrum is $X_\mathrm{mel}(k', b)$, where k' is the mel scale frequency bin. Then, the MFCCs are calculated by computing the discrete cosine transform (DCT) of the logarithm of the mapped power spectrum $X_\mathrm{mfcc}(\eta, b) = \mathrm{DCT}\{\log(X_\mathrm{mel}(k', b))\}$ with the cepstral coefficient index η. Again, to consider the temporal evolution, we compute the MFCC modulation spectrum $\hat{X}_\mathrm{mfcc}(\eta, \nu, c)$ similar to (1). Then, the absolute value of this modulation spectrum is averaged over all C_T frames and normalized to the zeroth modulation frequency band, similar to (2) and (3). The result is a small feature set, therefore no polynomial approximation is necessary.

2.1.3. Cepstral Normalization

Room dependent reverberation and additive noise are critical issues in the context of audio signal classification as they disturb audio signals in nearly every realistic situation. Reverberation is affected by changes of the source-receiver setup and by the room properties. The transmission path from a source to a receiver can be represented by the room impulse response (RIR). One approach to reduce the effect of reverberation in audio signal processing is the cepstral mean normalization (CMN). It is based on the idea that a convolutional distortion in the time domain corresponds to an additive term in the cepstral domain. By averaging over a certain amount of time, in which the RIR can be assumed to be constant and subtracting this average from the cepstrum, the influence of reverberation might be reduced [10] [11].

2.2. Signal clustering and classification

The goal of signal clustering is to assign objects to groups with small intra-group differences and large inter-group differences. In case of a clustering into groups without training data, unsupervised classification algorithms can be used to generate unlabeled clusters of objects [12]. In our investigations we use fuzzy clustering for partitioning clusters which are dominated by one of the sources. A fuzzy clustering algorithm, rather than making a hard-decision about the membership of an object to a specific cluster, estimates a probability $\mu = [0, 1]$, indicating the chance of an object to belong to each of the possible clusters [13]. This estimation may be done by minimizing an objective function which considers a μ-weighted distance between all objects and the estimated cluster centers. A-priori information about the number of clusters can be introduced to the algorithm or may be estimated as well. Fuzzy clustering is interesting for our work, as all microphones receive mixtures of all source signals. If microphones are close to a source, a high cluster membership probability can be assumed. For microphones with a balanced mixture of source signals and a rather high amount of reverberation, a smaller membership probability may be assumed, which offers the opportunity to exclude these microphones.

For the classification of specific classes we will utilize labeled data to train a classification system. Then, test data are assigned to the matching class [12]. In our investigation we used a linear discriminant analysis (LDA), which assumes a multivariate normal distribution for the feature vectors and a pooled estimate of the feature covariance matrix across all classes [14].

3. EXPERIMENT DESCRIPTION

For all experiments we simulated the scenario of ten microphones and two active target sources in a room of the size $6.7 \times 4.9 \times 3.5\mathrm{m}^3$ with a reverberation time $\mathrm{T}_{60} \approx 400\mathrm{ms}$ by creating RIRs using the method in [15] which provides realistic reverberation effects. To generate the microphone signals the two source signals were convolved with the respective RIRs and summed up. In this way each

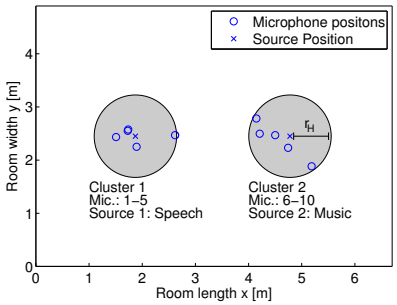

Fig. 2. The sources are placed at fixed positions in the room. The 10 microphones are split into 2 clusters and randomly positioned within the critical distance of each source.

microphone picked up signals from source 1 (clean speech, male and female, English, [16]) and source 2 (music, different genres, private database) (Fig. 2). To simulate a more realistic classification experiment an additional background noise class was added in the training step. For this third class different types of indoor noise sounds (e.g. vacuum cleaner, dish washer, private database) were used. However, these noise signals were not added to the microphone signals and thus were not part of the test data. The position of the sources was fixed to the approximate critical distance [17] $r_{\mathrm{H}} + 1$m alongside the wall in x- and half of the room size in y- and z-direction. The positions of the virtual microphones were randomized in x- and y-directions within the critical distance of a source, to generate distinct clusters. This positioning is motivated by the idea that within the critical distance the direct sound component is dominant. Thus, this arrangement of microphones allowed for a reasonable evaluation of the unsupervised clustering and the supervised classification experiments. However, the positioning information was not used in the following investigations as they were based purely on the extracted feature vectors.

For all investigations, we extracted LP-CMRARE and Mod-MFCC features, both with and without CMN, for signals of $T = 4$ seconds duration sampled at $f_s = 16$ kHz. For the spectral and cepstral analysis, the frame length was $N = 512$ and frame shift was $P = 256$ samples. For the cepstral modulation analysis the frame length and shift were set to $M = 16$ and $Q = 8$. We approximated the CMRs $r_{1|1}$ and $r_{2|8}$ using Legendre polynomial based parametrization of order $p = 12$, resulting in 26 coefficients. In case of Mod-MFCC features the normalized averaged modulation content was computed for 13 MFCCs, yielding again the modulation ratios $r_{1|1}$ and $r_{2|8}$ and thus resulting in 26 coefficients to summarize 4 seconds of data as well.

3.1. Unsupervised clustering of simulated microphone signals

To evaluate the ability of each feature set to form microphone clusters related to the dominant source contribution the *fuzzy c-means* algorithm of a freely available Matlab toolbox for fuzzy clustering was used [18]. As a priori information, the number of clusters was set to 2. The unsupervised clustering was evaluated with 50 scenarios of five microphones randomly positioned within the critical distance of each source, and 100 randomized combinations of speech (Source 1) and music (Source 2) signals for each scenario.

3.2. Supervised Classification of simulated microphone signals

To classify the microphones in the supervised classification problem we now considered the fixed scenario shown in Fig. 2. We used 100 audio files for each, speech, music and noise out of which 75% were used as clean data for training the LDA. The classification accuracy based on the training data was high for all feature sets (LP-CMRARE: 98.8%, LP-CMRARE CMN: 95.4%, Mod-MFCC: 96.2%, Mod-MFCC CMN: 93.5%). The remaining 25% of the data were used as test samples to simulate the microphone signals. The classification accuracy of each feature set was averaged over 50 cross-validation iterations, in which the combinations of speech and music and the allocation of test- and training data were randomized.

Table 1. Confusion matrices for the unsupervised microphone clustering task in % using the proposed feature sets.

	LP-CMRARE		LP-CMRARE CMN	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Mic. 1-5	89.2	10.8	90.9	9.1
Mic. 6-10	8.7	91.3	5.4	94.6

	Mod-MFCC		Mod-MFCC CMN	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Mic. 1-5	94.0	6.0	97.1	2.9
Mic. 6-10	3.2	96.8	3.4	96.6

3.3. Combined Experiment

In the combined experiment we trained the LDA classifier using 75% of each of the 100 clean and anechoic speech, music and noise samples, again for the scenario shown in Fig. 2. The remaining 25% of the speech and music data were processed (microphone simulation, feature extraction) and used for the unsupervised clustering, resulting in 25 cluster estimates. We only considered a microphone to be a valid member of one of the clusters, if the according probability provided by the fuzzy clustering algorithm was larger than 0.7. Thereby, we excluded signals with a low SDR. Finally, we computed a new feature vector per cluster using weighted averages of the feature vectors obtained by the cluster microphones. The weighting factor accorded to the membership probability of a microphone for a cluster. This new vector was used as test instance for the LDA. To rate the performance of this weighted feature vector, we also generated a non-weighted averaged feature vector based on all five microphones within the critical distance of one source as test data for the LDA for each of the 25 convolved mixtures. Again, we performed a 50-fold cross-validation.

4. RESULTS

4.1. Unsupervised clustering of simulated microphone signals

Table 1 provides the averaged success rates of assigning microphones 1-5 to cluster 1 and microphones 6-10 to cluster 2 using the unsupervised fuzzy clustering algorithm. All feature sets provided distinct characteristics for a successful clustering decision. By applying CMN, the results improve slightly. The highest accuracy is provided by the Mod-MFCC with CMN.

4.2. Supervised Classification of simulated microphone signals

Table 2 presents the classification results in % for all simulated microphone signals and the fixed scenario shown in Fig. 2. Microphones 1-5 are located in cluster 1 and should provide the classification result speech, whereas microphones 6-10 in cluster 2 should provide the classification result music. Both features, LP-CMRARE and Mod-MFCC, highly misclassify speech when cepstral normalization is not used. As soon as CMN reduces the effects of reverberation, the results for speech detection improve. In those cases, a

Table 2. Averaged classification results in % for LDA classification of single microphone signals in a fixed source-microphone scenario.

		Cluster 1 (desired result: Speech)					Cluster 2 (desired result: Music)				
	Mic.-No.	1	2	3	4	5	6	7	8	9	10
LP-CMRARE	Speech	33.9	19.5	2.3	45.2	**50.3**	0.0	1.8	0.8	0.1	0.4
	Music	**66.1**	**80.5**	**97.7**	**54.8**	49.7	**99.9**	**97.7**	**98.0**	**99.7**	**99.2**
	Noise	0.0	0.0	0.0	0.0	0.0	0.1	0.5	1.2	0.2	0.4
LP-CMRARE	Speech	**69.3**	45.4	12.0	**70.5**	**78.0**	1.8	1.7	1.0	3.3	1.3
CMN	Music	30.6	**54.6**	**88.0**	29.4	22.0	**97.0**	**91.3**	**90.3**	**92.8**	**96.6**
	Noise	0.1	0.0	0.0	0.1	0.0	1.2	7.0	8.7	3.9	2.1
Mod-MFCC	Speech	6.4	0.0	0.0	6.4	10.5	0.0	0.0	0.3	0.0	0.0
	Music	**93.6**	**100.0**	**100.0**	**93.6**	**89.5**	**99.9**	**99.4**	**99.0**	**100.0**	**99.8**
	Noise	0.0	0.0	0.0	0.0	0.0	0.1	0.6	0.7	0.0	0.2
Mod-MFCC	Speech	**58.5**	31.8	6.0	**59.3**	**62.6**	3.0	6.0	6.3	3.8	3.0
CMN	Music	41.5	**68.2**	**94.0**	40.7	37.4	**96.4**	**91.4**	**91.1**	**96.0**	**96.5**
	Noise	0.0	0.0	0.0	0.0	0.0	0.6	2.6	2.6	0.2	0.5

Table 3. Averaged classification results in % for LDA classification using averaged feature vectors of all microphones in one cluster and fuzzy-weighted averaged feature vectors.

		Microphones within critical distance		Fuzzy Cluster Estimation	
		Cl. 1	Cl. 2	Cl. 1	Cl. 2
LP-CMRARE	Speech	19.7	0.1	40.3	4.0
	Music	**80.3**	**99.8**	**59.7**	**96.0**
	Noise	0.0	0.1	0.0	0.0
LP-CMRARE	Speech	**57.4**	1.5	**92.0**	16.0
CMN	Music	42.6	**94.8**	8.0	**80.0**
	Noise	0.0	3.7	0.0	4.0
Mod-MFCC	Speech	1.1	0.0	8.0	0.0
	Music	**98.9**	**100.0**	**92.0**	**100.0**
	Noise	0.0	0.0	0.0	0.0
Mod-MFCC	Speech	43.2	2.1	44.0	0.0
CMN	Music	**56.8**	**96.9**	**56.0**	**100.0**
	Noise	0.0	1.0	0.0	0.0

majority decision would deliver the desired result. The poor classification result of microphone 3 for all feature sets in this example is related to the position of the microphone at the very right of cluster 1 (Fig. 2), which has the consequence of the lowest SDR of all microphone signals in cluster 1. Music was very well classified for all microphones in cluster 2. Misclassification as noise only occurs for LP-CMRARE with CMN to a noticeable amount.

4.3. Combined Experiments

Table 3 shows the results for averaged feature vectors. Results shown in the first two columns are based on feature vectors equally averaged over all five microphones in each cluster. Again, speech classification is performed with a poor classification result and applying CMN improves the result. For the results presented in the last two columns the cluster memberships probabilities estimated by the fuzzy algorithm were used. Here, a feature vector is considered for the weighted averaging only in the case of a high cluster membership probability (> 0.7). Therefore, microphones with a low SDR might be excluded here. The result for speech classification in cluster 1 for the LP-CMRARE based feature vector improves, especially in combination with CMN. For the Mod-MFCC feature vectors the weighted averaging does not improve the classification results to the same extend. In cluster 2 music is almost always recognized. Here, the LP-CMRARE CMN have the lowest classification rate with 80%. Misclassification of one of the clusters as background noise occurs just very rarely.

5. DISCUSSION

The unsupervised fuzzy clustering of the microphones in two clusters works very well for a combination of a speech source and a music source in the simulated scenarios. Interestingly, the CMN brings only a slight improvement. All feature vectors seem to be distinguishable despite the reverberation. This can be explained by the fact that only distorted features are compared and no clean reference is used. The classification experiments show that the correct classification of speech signals is a difficult problem when reverberation and a competing music source are present. The observed shift (e.g. Tab. 3) towards the classification result music may be related to natural speech pauses in which however music is present in the microphone signals. The reduction of reverberation effects using CMN improved the single channel classification results, thus a majority decision in a cluster might deliver the desired result .

The approach of generating a new feature vector by averaging all five feature vectors in a cluster to obtain one classification result per cluster delivers comparable results to a majority decision of single microphone classification results within a cluster. The exploitation of cluster membership probabilities delivered by the fuzzy clustering for the generation of a smoothed feature vector works well for LP-CMRARE CMN features and improves the correct classification of speech in these cases.

6. CONCLUSION

The performance of audio signal classification is reduced drastically when environmental conditions influence the test data and therefore lead to a mismatch between training and test data. Our investigation showed that ad-hoc microphone arrays, cepstral audio features and fuzzy clustering can be used to obtain solid cluster-based classification results for simulated reverberant audio signals in a two-source scenario, although the training data for the classifier was exclusively clean and undistorted data. The amount of data to transmit from array components to a central clustering and classification unit is relatively small as only 26 coefficients for 4 seconds of analysis time are necessary. In a future investigation more flexible source-receiver-setups in different reverberant environments and an evaluation on real recorded data will be tackled.

7. ACKNOWLEDGMENT

The authors would like to thank Prof. D. Kolossa for valuable comments on the application of CMN.

8. REFERENCES

[1] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 2915 – 2929, 2005.

[2] A. R. Abu-El-Quran, R. A. Goubran, and A. D. C. Chan, "Security monitoring using microphone arrays and audio classification.," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1025–1032, 2006.

[3] I. Himawan, I. McCowan, and S. Sridharan, "Clustering of ad-hoc microphone arrays for robust blind beamforming," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[4] M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007.

[5] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen, "Closed-form self-localization of asynchronous microphone arrays," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011.

[6] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011.

[7] R. Martin and A. Nagathil, "Cepstral modulation ratio regression (CMRARE) parameters for audio signal analysis and classification," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009.

[8] A. Nagathil, P. Göttel, and R. Martin, "Hierarchical audio classification using cepstral modulation ratio regressions based on legendre polynomials," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011.

[9] S. Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., 2000.

[10] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García D., Petrovska-Delacrétaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 430–451, Jan. 2004.

[11] P. N. Garner, "Cepstral normalisation and the signal to noise ratio spectrum in automatic speech recognition," *Speech Communication*, vol. 53, no. 8, pp. 991–1001, October 2011.

[12] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[13] M. S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer Modelling*, vol. 18, pp. 1–16, 1993.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, 2. edition, 2001.

[15] S. Gergen, C. Borß, N. Madhu, and R. Martin, "An optimized parametric model for the simulation of reverberant microphone signals," in *Proc. of the International Conference on Signal Processing, Communications and Computing (ICSPCC 2012)*, 2012.

[16] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus," 1993, Linguistic Data Consortium, Philadelphia.

[17] H. Kuttruff, *Room Acoustics*, Applied Science Publishers Ltd, London, 1979.

[18] J. Abonyi, "Fuzzy clustering and data analysis toolbox," April 2005.