Neural Process Lett (2019) 49:1299–1320
https://doi.org/10.1007/s11063-018-9868-7

Bounded Generalized Gaussian Mixture Model with ICA

Muhammad Azam[1] · **Nizar Bouguila**[2]

Published online: 25 June 2018

Abstract In this paper, we propose bounded generalized Gaussian mixture model with independent component analysis (ICA). One limitation in ICA is that it assumes the sources to be independent from each other. This assumption can be relaxed by employing a mixture model. In our proposed model, bounded generalized Gaussian distribution (BGGD) is adopted for modeling the data and we have further extended its mixture as an ICA mixture model by employing gradient ascent along with expectation maximization for parameter estimation. By inferring the shape parameter in BGGD, Gaussian distribution and Laplace distribution can be characterized as special cases. In order to validate the effectiveness of this algorithm, experiments are performed on blind source separation (BSS) and BSS as preprocessing to unsupervised keyword spotting. For BSS, TIMIT, TSP and Noizeus speech corpora are selected and results are compared with ICA. For keyword spotting, TIMIT speech corpus is selected and recognition results are further compared before and after BSS being applied as preprocessing when speech utterances are affected by mixing of noise or other speech utterances. The mixing of noise or speech utterances with a particular or target speech utterance can greatly affect the intelligibility of a speech signal. The results achieved from the presented experiments on different applications have demonstrated the effectiveness of ICA mixture model in statistical learning.

Keywords Bounded generalized Gaussian mixture model (BGGMM) · Independent component analysis (ICA) · Blind source separation (BSS) · Unsupervised keyword spotting · Segmental dynamic time warping (DTW)

✉ Muhammad Azam
mu_azam@encs.concordia.ca

Nizar Bouguila
nizar.bouguila@concordia.ca

1 Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

2 Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

1 Introduction

In machine learning and pattern recognition, effectiveness of an approach or an algorithm is determined by the ability of modeling underlying distribution of observed data [38]. Finite mixture models have been extensively used for statistical modeling in machine learning and pattern recognition and have demonstrated their importance in many speech and image processing applications [33,60]. Gaussian mixture model (GMM) is well renowned for data clustering. The parameters of GMM can be estimated effectively using expectation maximization (EM) algorithm by maximizing the log-likelihood function [10,51]. The main problem associated with GMM is sensitivity to outliers [51]. Student's-t mixture model (SMM) has been proposed in order to improve the robustness of Gaussian mixture model for statistical modeling [44,56,76]. In SMM, each component has one more parameter, called degree of freedom, as compared to GMM. Cauchy and Gaussian distributions are special cases of student's-t distribution with degree of freedom 1 and ∞, respectively [51]. There have been substantial growth in research for developing mixture models using generalized Gaussian distribution (GGD) [2–4,13,45]. This distribution has one extra parameter (shape parameter λ) than Gaussian distribution, which controls the tails of distribution. One problem associated with above mentioned mixture models is unbounded support range $(-\infty,+\infty)$ of their distributions [51]. It is observed that many real applications have their data within bounded support regions [21,26,43]. For speech processing applications, bounded Gaussian mixture model (BGMM) has been proposed in [26,43]. The idea of bounded support mixture is adopted for GGMM and BGGMM has been proposed in [51], which provides a generalization for GMM, Laplace mixture model (LMM), GGMM and BGMM as special cases.

ICA mixture model has been proposed as an extension of Gaussian mixture model in [38,39,41]. ICA has been successfully applied to problems such as blind source separation and signal analysis describing its ability to model non-Gaussian statistical structures. If the source distributions are assumed to be Gaussian, it is equivalent to principle component analysis (PCA), which assumes that observed data is distributed as a multivariate Gaussian [38]. ICA generalize PCA by modeling the observed data with non-Gaussian distributions and goal is to linearly transform the data structures in such a way that variables after transformation are independent from each other [41]. One limitation in ICA is that it assumes the sources to be independent from each other. This assumption can be relaxed by employing a mixture model. The observed data can be categorized into several mutually exclusive classes by employing a mixture model [40], simply called an ICA mixture model. It can be generalized with the assumption that observed data in each class is produced by a linear combination of independent, non-Gaussian sources as in case of ICA [41]. Hence, in an ICA mixture model, it is assumed that observed data can be categorized into mutually exclusive classes and components of the model are generated by linear combination of independent sources [66]. Many variations of ICA mixture model have been proposed in the last few years [12, 48,67,68]. It has been extensively used for statistical modeling in a variety of applications that include segmentation, image enhancement and BSS [39,41,62]. In [52], ICA mixture model was proposed with adaptive source densities including generalized Gaussian and Student's t distributions as special cases along with other forms of densities. In this paper, we are interested in extending the model presented in [38] with BGGD. In [51], BGGMM is formulated for univariate data which is extended here for multivariate data. The parameter estimation for proposed ICA mixture is adopted from [38,41] using ICA and gradient ascent. The preliminary results obtained by applying the proposed ICA mixture are published in [6,7].

In this paper we have extended the applications of ICA mixture in BSS and unsupervised keyword spotting frameworks for more insightful analysis.

Blind source separation has been applied to many signal processing and machine learning problems including speech enhancement, speech recognition, medical signal processing and telecommunications [41]. BSS is defined as a method which reconstruct the unknown sources of observed signals from an unknown mixture [11,57,72,73]. BSS was formulated around 1982 and first related contributions appeared around 1985 in [5,16,27–29,35]. The ICA was proposed as general framework for solving blind source separation problems based on statistical independence of the unknown sources in [34] and formalized for linear mixtures in [14,15]. The limitations associated with ICA were controlled by ICA mixture as proposed in [37,38] and successfully applied to BSS [8,40,41]. Research for the development of many new approaches for BSS is continued and many interesting algorithms and techniques have been developed [1,24]. The Expectation-Maximization (EM) algorithm has also been applied to ICA in [32,59]. In this paper, we have proposed BGGMM using ICA for the task of BSS. For the evaluation of proposed BSS framework, we have used signal-to-distortion ratio (SDR), signal-to-interference ratio (SIR), signal-to-artifact ratio (SAR) and perceptual evaluation of speech quality (PESQ). The detailed explanation of evaluation metrics is presented in [20,49,58,74].

Automatic speech recognition (ASR) is considered as a nonlinear transformation from spoken words to text [61,79], which requires large quantities of annotated data along with the language specific speech and text data, used for training complex statistical acoustic and language models [31,78,80]. Keyword spotting task has also been explored for many years and ASR is used to detect the occurrence of a specific keyword in speech data [70]. Keyword spotting is defined as an approach for speech understanding to detect specific keyword(s) that most likely express the intent of a speaker rather than recognition of a whole speech utterance [46]. Hidden Markov models based keyword spotting methods have been proposed widely for supervised and unsupervised settings [63,64,69,71,77]. Dynamic time warping has been used extensively for speech recognition and keyword spotting [17,50,53–55,65,81,82]. The use of mixture model in automatic speech recognition and keyword spotting has demonstrated its effectiveness in unsupervised platforms and settings [42,75]. We have proposed BGGMM with ICA in unsupervised keyword spotting and preliminary results are submitted in [6]. In many real time scenarios, speech signals are mixed with noise or other speech signal which reduces the intelligibility of signals in keyword spotting and speech recognition. In order to improve the detection rate in keyword spotting, speech signal can be pre-processed using BSS before being applied to the trained model for keyword detection or speech recognition. The proposed ICA mixture have demonstrated its effectiveness in BSS as described in Sect. 3.1 and we have proposed the same BSS framework as prepossessing to unsupervised keyword spotting presented in [6]. Due to mixing of speech utterances, two types of problems occur in keyword spotting. In the first case, target keyword will more likely not detected during the keyword spotting, whereas in second case target keyword will be detected in correct speech utterance but it will also get detected in other speech utterances as false alarm. These two problems are explained in detail in Sect. 3.2. In this paper, we have proposed BSS as pre-processing to unsupervised keyword spotting as an extension to the work submitted in [6].

The rest of the paper is organized as follows. Section 2 describes derivation of learning rules for proposed algorithm. In Sect. 3, application of ICA mixture model in BSS and unsupervised keyword spotting is presented with set of experiments and results. Section 4 presents the conclusions and future perspectives.

2 Bounded Generalized Gaussian Mixture Model with ICA

In this section, BGGMM with ICA is proposed for statistical modeling. In an ICA mixture model, it is assumed that observed data come from a mixture model and it can be categorized into mutually exclusive classes which means that each class of the data is modeled via an ICA [38,66]. Consider the case where input is a set of features of the data represented as $\mathcal{X} = (\vec{X}_1, \ldots, \vec{X}_N)$ and $\vec{X}_i$ is a D-dimensional random variable $\vec{X}_i = [X_{i1}, \ldots, X_{iD}]^T$. The $\vec{X}_i$ follows a K components mixture distribution if its probability function can be written as follows:

$$p(\vec{X}_i|\Theta) = \sum_{j=1}^{K} p(\vec{X}_i|\xi_j)p_j \tag{1}$$

provided that $p_j \geq 0$ and $\sum_{j=1}^{K} p_j = 1$. In Eq. (1), $p(\vec{X}_i|\xi_j)$ is probability density function, ξ_j represents the set of parameters defining jth component, p_j is mixing proportion, $\Theta = (\xi_1, \ldots, \xi_K, p_1, \ldots, p_K)$ is complete set of parameters to characterize the mixture model and $K \geq 1$ is number of components in the mixture model [18,22,47]. For an ICA mixture model, each data vector $\vec{X}_i$ can be represented as:

$$\vec{X}_i = \mathrm{A}_j \vec{\mathrm{s}}_{j,i} + \vec{\mathrm{b}}_j \tag{2}$$

where A_j is $L \times D$ scalar matrix termed as basis functions, $\vec{\mathrm{s}}_{j,i}$ is D-dimensional source vector and $\vec{\mathrm{b}}_j$ is an L-dimensional bias vector for a particular mixture component j [38,39,41,66–68]. In order to define the BGGD for a variable $\vec{X} \in \mathbb{R}$, it is required to provide an indicator function which introduces the boundary conditions. For each component (denoted by j), indicator function $\mathrm{H}(\vec{X}_i|j)$ is defined with bounded support region (Ω_{S_j}) for each component:

$$\mathrm{H}(\vec{X}_i|j) = \begin{cases} 1 & \text{if } \vec{X}_i \in \Omega_{S_j} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

For BGGMM, $\vec{X}_i$ follows a K components mixture represented in Eq. (1), where $p(\vec{X}_i|\xi_j)$ is multivariate BGGD as:

$$p(\vec{X}_i|\xi_j) = \frac{f_{ggd}(\vec{X}_i|\xi_j)\mathrm{H}(\vec{X}_i|j)}{\int_{\Omega_{S_j}} f_{ggd}(\vec{\mathrm{u}}|\xi_j)d\mathrm{u}} \tag{4}$$

where term $f_{ggd}(\vec{X}_i|\xi_j)$ represents the multivariate generalized Gaussian distribution (GGD):

$$f_{ggd}(\vec{X}_i|\xi_j) = \prod_{d=1}^{D} \frac{\lambda_{jd}\sqrt{\frac{\Gamma(3/\lambda_{jd})}{\Gamma(1/\lambda_{jd})}}}{2\sigma_{jd}\Gamma(1/\lambda_{jd})} \exp\left(-A(\lambda_{jd})\left|\frac{X_{id} - \mu_{jd}}{\sigma_{jd}}\right|^{\lambda_{jd}}\right) \tag{5}$$

with

$$A(\lambda_{jd}) = \left[\frac{\Gamma(3/\lambda_{jd})}{\Gamma(1/\lambda_{jd})}\right]^{\lambda_{jd}/2} \tag{6}$$

The term $\int_{\Omega_{S_j}} f_{ggd}(\vec{\mathrm{u}}|\xi_j)d\mathrm{u}$ is normalization constant that indicates the share of $f_{ggd}(\vec{X}_i|\xi_j)$ which belongs to the support region. Note that $\xi_j = \left\{\vec{\mu}_j, \vec{\sigma}_j, \vec{\lambda}_j, \mathrm{A}_j, \vec{\mathrm{b}}_j\right\}$ is the set of parameters defining jth component, where $\vec{\mu}_j = (\mu_{j1}, \ldots, \mu_{jD})$, $\vec{\sigma}_j = (\sigma_{j1}, \ldots, \sigma_{jD})$,

$\vec{\lambda}_j = (\lambda_{j1}, \ldots, \lambda_{jD})$, $\mathrm{A}_j = (\mathrm{a}_1, \ldots, \mathrm{a}_L)$ and $\vec{\mathrm{b}}_j = (\mathrm{b}_{j1}, \ldots, \mathrm{b}_{jD})$ are the mean, standard deviation, shape parameters, basis functions and bias vector, respectively. The vectors representing mean, standard deviation, shape parameters and bias are D-dimensional for each component of the mixture model, whereas the basis functions for each component has L number of linear combination with each linear combination being D-dimensional. For simplicity, number of linear combinations (L) is considered to be equal to the number of sources (D) in each observation which makes basis functions a $D \times D$ scalar matrix. With a mixture of K BGGDs, the likelihood of data $\mathcal{X}$ can be defined as:

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} p(\vec{X}_i|\xi_j) p_j \tag{7}$$

where complete set of parameters of the ICA mixture model having K classes is defined by $\Theta = (\vec{\mu}_1, \ldots, \vec{\mu}_K, \vec{\sigma}_1, \ldots, \vec{\sigma}_K, \vec{\lambda}_1, \ldots, \vec{\lambda}_K, \mathrm{A}_1, \ldots, \mathrm{A}_K, \vec{\mathrm{b}}_1, \ldots, \vec{\mathrm{b}}_K, p_1, \ldots, p_K)$. We introduce the stochastic indicator $\mathcal{Z} = \{\vec{Z}_1, \ldots, \vec{Z}_N\}$, where $\vec{Z}_i = (Z_{i1}, \ldots, Z_{iK})$ is the label of each observation, such that $Z_{ij} \in \{0, 1\}$, $\sum_{j=1}^{K} Z_{ij} = 1$. The role of these variables is to encode membership of each observation for a relative component of the mixture model. In other words, Z_{ij}, the unobserved variable in each indicator vector equals to 1 if $\vec{X}_i$ belongs to class j and 0, otherwise [10,18,19]. The complete data likelihood is:

$$p(\mathcal{X}, \mathcal{Z}|\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{K} \left(p(\vec{X}_i|\xi_j) p_j \right)^{Z_{ij}} \tag{8}$$

For instance, if we consider that number of mixture components is known, then parameter estimation requires the maximization of log-likelihood function:

$$\mathcal{L}(\Theta, \mathcal{Z}, \mathcal{X}) = \sum_{i=1}^{N} \sum_{j=1}^{K} Z_{ij} \log \left(p(\vec{X}_i|\xi_j) p_j \right) \tag{9}$$

By replacing each Z_{ij} with its expectation, defined as posterior probability that the ith observation belongs to jth component of the mixture model we obtain:

$$\hat{Z}_{ij} = p(j|\vec{X}_i) = \frac{p(\vec{X}_i|\xi_j) p_j}{\sum_{j=1}^{K} p(\vec{X}_i|\xi_j) p_j} \tag{10}$$

2.1 Parameters Estimation

In a mixture model, parameters include mixing proportions and parameters of the distribution whereas in case of ICA mixture model each vector of the data is represented as in Eq. (2), which also necessitates the estimation of basis functions and bias vectors. The basis functions and bias vectors are further adopted to compute the sources in ICA model. For the parameters mean, standard deviation and mixing proportions, maximization of log-likelihood is obtained by setting the gradient of log-likelihood (with respect to each parameter) to zero. The maximization of log-likelihood for the shape parameters, basis functions and bias vector is performed by employing standard ICA model and gradient ascent. Using Eq. (10), each observation can be labeled to one or zero for a particular component of the mixture model which can be further applied to maximize the complete data log-likelihood with respect to parameters of ICA mixture model. The gradient of log-likelihood with respect to parameters of each component is computed as following:

$$\nabla_{\Theta_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X}) = \nabla_{\Theta_j}\sum_{i=1}^{N}\sum_{j=1}^{K} Z_{ij}\log\left(p(\vec{X}_i|\xi_j)p_j\right) \tag{11}$$

The ∇_{Θ_j} represents here the gradient with respect to p_j, $\vec{\mu}_j$, $\vec{\sigma}_j$, $\vec{\lambda}_j$, A_j and $\vec{\mathrm{b}}_j$. Eq. (11) can be written as:

$$\begin{aligned}&\nabla_{\Theta_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X})\\ &= \nabla_{\Theta_j}\sum_{i=1}^{N}\sum_{j=1}^{K} Z_{ij}\times\left\{\log p_j + \log f_{ggd}(\vec{X}_i|\xi_j) + \log \mathrm{H}(\vec{X}_i|j) - \log\int_{\Omega_{S_j}} f_{ggd}(\vec{\mathrm{u}}|\xi_j)d\mathrm{u}\right\}\end{aligned} \tag{12}$$

2.1.1 Estimation of Mixing Parameter, Mean and Standard Deviation

The mixing parameter can be estimated by taking the gradient of complete data log-likelihood with respect to p_j. In order to ensure the constraints $p_j > 0$ and $\sum_{j=1}^{M} p_j = 1$, a Lagrange multiplier is introduced while estimating p_j. Thus, the augmented log-likelihood function can be expressed by:

$$\Phi(\Theta,\mathcal{Z},\mathcal{X},\Lambda) = \sum_{i=1}^{N}\sum_{j=1}^{K} Z_{ij}\log\left(p(\vec{X}_i|\xi_j)p_j\right) + \Lambda\left(1-\sum_{j=1}^{K} p_j\right) \tag{13}$$

where Λ is the Lagrange multiplier. Taking the gradient of augmented log-likelihood function with respect to p_j, we get the estimated value of p_j as:

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N} p(j|\vec{X}_i) \tag{14}$$

where $p(j|\vec{X}_i)$ is the posterior probability. The mean μ_j can be estimated by maximizing the log-likelihood with respect to μ_j. The estimated mean $\hat{\mu}_{jd}$ for $d = 1,\ldots,D$ is given by:

$$\begin{aligned}\hat{\mu}_{jd} = \frac{1}{\sum_{i=1}^{N}\hat{Z}_{ij}\left|X_{id}-\mu_{jd}\right|^{(\lambda_{jd}-2)}}\sum_{i=1}^{N}\hat{Z}_{ij}\Bigg\{&\left[\left|X_{id}-\mu_{jd}\right|^{(\lambda_{jd}-2)}X_{id}\right]\\ &-\left[\frac{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)\mathrm{sign}\left(\mathrm{u}-\mu_{jd}\right)\left|\mathrm{u}-\mu_{jd}\right|^{\lambda_{jd}-1}d\mathrm{u}}{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)d\mathrm{u}}\right]\Bigg\}\end{aligned} \tag{15}$$

Note that, in Eq. (15), the term $\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)\mathrm{sign}\left(\mathrm{u}-\mu_{jd}\right)\left|\mathrm{u}-\mu_{jd}\right|^{\lambda_{jd}-1}d\mathrm{u}$ is the expectation of function $\mathrm{sign}\left(\mathrm{u}-\mu_{jd}\right)\left|\mathrm{u}-\mu_{jd}\right|^{\lambda_{jd}-1}$ under the probability distribution $f_{ggd}(\mathrm{u}|\xi_j)$ [10,21,51], which can be approximated as:

$$\begin{aligned}&\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)\mathrm{sign}\left(\mathrm{u}-\mu_{jd}\right)\left|\mathrm{u}-\mu_{jd}\right|^{\lambda_{jd}-1}d\mathrm{u}\\ &\approx \frac{1}{M}\sum_{m=1}^{M}\mathrm{sign}(\mu_{jd}-s_{j_{md}})\left|\mu_{jd}-s_{j_{md}}\right|^{\lambda_{jd}-1}\mathrm{H}(s_{j_{md}}|j)\end{aligned} \tag{16}$$

where $s_{m_{jd}} \sim f_{ggd}(\mathrm{u}|\xi_j)$ is a set of random variables drawn from generalized Gaussian distribution for particular component of the mixture model j. The set of data with random variables have M vectors with D dimensions. M is a large integer chosen to generate the set of random variables. Similarly, the term $\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)d\mathrm{u}$ in Eq. (15) can be approximated as:

$$\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)d\mathrm{u} \approx \frac{1}{M}\sum_{m=1}^{M} \mathrm{H}(s_{m_{jd}}|j) \tag{17}$$

From Eqs. (16) and (17), $\hat{\mu_j}$ can be written as:

$$\hat{\mu}_{jd} = \frac{1}{\sum_{i=1}^{N}\hat{Z}_{ij}\left|X_{id}-\mu_{jd}\right|^{(\lambda_{jd}-2)}} \sum_{i=1}^{N}\hat{Z}_{ij}\left\{\left[\left|X_{id}-\mu_{jd}\right|^{(\lambda_{jd}-2)}X_{id}\right] - \left[\frac{\sum_{m=1}^{M}\mathrm{sign}(\mu_{jd}-s_{j_{md}})\left|\mu_{jd}-s_{j_{md}}\right|^{\lambda_{jd}-1}\mathrm{H}(s_{j_{md}}|j)}{\sum_{m=1}^{M}\mathrm{H}(s_{j_{md}}|j)}\right]\right\} \tag{18}$$

with $i = 1, \ldots, N$, $j = 1, \ldots, K$, $d = 1, \ldots, D$ and $m = 1, \ldots, M$. The standard deviation σ_j can be estimated by maximizing the log-likelihood with respect to σ_j. The estimated standard deviation $\hat{\sigma}_{jd}$ for $d = 1, \ldots, D$ is given as:

$$\hat{\sigma}_{jd} = \left(\frac{\sum_{i=1}^{N}\hat{Z}_{ij}\left[A(\lambda_{jd})\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}}\lambda_{jd}\right]}{\sum_{i=1}^{N}\hat{Z}_{ij}\left\{1+\left[\frac{\int_{\partial\Omega_j} f_{ggd}(\mathrm{u}|\xi_j)\left\{-1+A(\lambda_{jd})\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}}\lambda_{jd}(\sigma_{jd})^{-\lambda_{jd}}\right\}d\mathrm{u}}{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)d\mathrm{u}}\right]\right\}}\right)^{1/\lambda_{jd}} \tag{19}$$

Similar to Eq. (16), in Eq. (19) the term $\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)(-1 + A(\lambda_{jd})\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}}\lambda_{jd}(\sigma_{jd})^{-\lambda_{jd}})d\mathrm{u}$ can be approximated as:

$$\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j)(-1 + A(\lambda_{jd})\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}}\lambda_{jd}(\sigma_{jd})^{-\lambda_{jd}})d\mathrm{u} \approx \frac{1}{M}\sum_{m=1}^{M}(-1+\lambda_{jd}A(\lambda_{jd})\left|s_{mjd}-\mu_{jd}\right|^{\lambda_{jd}}(\sigma_{jd})^{-\lambda_{jd}})\mathrm{H}(s_{mjd}|j) \tag{20}$$

From Eqs. (20) and (17), $\hat{\sigma_j}$ can be written as:

$$\hat{\sigma}_{jd} = \left(\frac{\sum_{i=1}^{N}\hat{Z}_{ij}\left[A(\lambda_{jd})\left|X_{id}-\mu_{jd}\right|^{\lambda_{jd}}\lambda_{jd}\right]}{\sum_{i=1}^{N}\hat{Z}_{ij}\left\{1+\left[\frac{\sum_{m=1}^{M}(-1+\lambda_{jd}A(\lambda_{jd})\left|s_{mjd}-\mu_{jd}\right|^{\lambda_{jd}}(\sigma_{jd})^{-\lambda_{jd}})\mathrm{H}(s_{mjd}|j)}{\sum_{m=1}^{M}\mathrm{H}(s_{mjd}|j)}\right]\right\}}\right)^{1/\lambda_{jd}} \tag{21}$$

with $i = 1, \ldots, N$, $j = 1, \ldots, K$, $d = 1, \ldots, D$ and $m = 1, \ldots, M$.

2.1.2 Parameter Estimation using ICA and Gradient Ascent

For parameter estimation using ICA and gradient ascent, zero mean and unit variance is assumed which is fundamental assumption of the source in ICA. The parameters estimated using ICA with gradient ascent include basis functions, bias vector and shape parameters. The gradient of complete data log-likelihood for the parameters of each class is given below:

$$\nabla_{\Theta_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X}) = \sum_{i=1}^{N}\sum_{j=1}^{K} p(j|\vec{X}_i)\nabla_{\Theta_j}\log\left(p(\vec{X}_i|\xi_j)p_j\right) \tag{22}$$

The ∇_{Θ_j} represents here the gradient with respect to basis function, bias vector and shape parameter.

$$\nabla_{\Theta_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X}) = \sum_{i=1}^{N}\sum_{j=1}^{K} p(j|\vec{X}_i)\left(\nabla_{\Theta_j}\log p(\vec{X}_i|\xi_j) + \nabla_{\Theta_j}\log p_j\right) \tag{23}$$

The term $\nabla_{\Theta_j}\log p_j$ will become zero while taking gradient with respect to basis functions, bias vector and shape parameter which will lead us to :

$$\nabla_{\Theta_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X}) = \sum_{i=1}^{N}\sum_{j=1}^{K} p(j|\vec{X}_i)\left(\nabla_{\Theta_j}\log p(\vec{X}_i|\xi_j)\right) \tag{24}$$

The class log-likelihood $\log p(\vec{X}_i|\xi_j)$ in Eq. (24) can be estimated using standard ICA model as follows:

$$\log p(\vec{X}_i|\xi_j) = \log\frac{p(\vec{\mathrm{s}}_{j,i})}{\left|\mathrm{detA}_j\right|} \tag{25}$$

The source can be computed by applying estimated basis function and bias vector in the above equation and log-likelihood of the standard ICA model will become:

$$\log p(\vec{X}_i|\xi_j) = \log p(\mathrm{A}_j{}^{-1}(\vec{X}_i - \vec{\mathrm{b}}_j)) - \log\left|\mathrm{detA}_j\right| \tag{26}$$

(a) *Basis Functions Estimation* The adaptation of basis functions for each component of ICA mixture is performed by maximizing the log-likelihood with respect to basis functions A_j for each component of mixture model:

$$\nabla_{\mathrm{A}_j}\mathcal{L}(\Theta,\mathcal{Z},\mathcal{X}) = \sum_{i=1}^{N} p(j|\vec{X}_i)\nabla_{\mathrm{A}_j}\log p(\vec{X}_i|\xi_j) \tag{27}$$

The adaptation performed by the gradient ascent with respect to basis functions is given as:

$$\Delta\mathrm{A}_j \propto p(j|\vec{X}_i)\frac{\partial}{\partial\mathrm{A}_j}\log p(\vec{X}_i|\xi_j) \tag{28}$$

The derivative in Eq. (28) can be computed using derivations given in standard ICA learning algorithm [41].

$$\frac{\partial}{\partial\mathrm{A}_j}\log p(\vec{X}_i|\xi_j) = \mathrm{A}_j\left[\mathrm{I} - 2\tanh(\vec{\mathrm{s}}_{j,i})\vec{\mathrm{s}}_{j,i}^{T}\right] \tag{29}$$

By using the standard ICA model for log-likelihood, we get:

$$\Delta\mathrm{A}_j \propto p(j|\vec{X}_i)\mathrm{A}_j\left[\mathrm{I} - 2\tanh(\vec{\mathrm{s}}_{j,i})\vec{\mathrm{s}}_{j,i}^{T}\right] \tag{30}$$

In adaptation of basis functions, the gradient of component of mixture model with respect to basis functions is weighted by $p(j|\vec{X}_i)$. An estimate of basis functions using gradient ascent is as follows:

$$\hat{\mathrm{A}}_j = \mathrm{A}_j + \alpha \left(p(j|\vec{X}_i)\mathrm{A}_j \left[\mathrm{I} - 2\tanh(\vec{\mathrm{s}}_{j,i})\vec{\mathrm{s}}_{j,i}^T \right] \right) \tag{31}$$

where α is step size and source is represented as:

$$\vec{\mathrm{s}}_{j,i} = {\mathrm{A}_j}^{-1}(\vec{X}_i - \vec{\mathrm{b}}_j) \tag{32}$$

(b) *Bias Vectors Estimation* The adaptation of bias vector can be performed for each component of the mixture model by using the Eq. (24).

$$\nabla_{\mathrm{b}_{jd}} \mathcal{L}(\Theta, \mathcal{Z}, \mathcal{X}) = \sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\mathrm{b}_{jd}} \log p(\vec{X}_i|\xi_j) \tag{33}$$

The gradient ascent is used for adaptation, with gradient of the component density with respect to bias term b_{jd} for each component of the mixture model:

$$\Delta \mathrm{b}_{jd} \propto p(j|\vec{X}_i) \frac{\partial}{\partial \mathrm{b}_{jd}} \log p(\vec{X}_i|\xi_j) \tag{34}$$

The Eq. (26) can be applied in Eq. (34) to adapt the bias term:

$$\Delta \mathrm{b}_{jd} \propto p(j|\vec{X}_i) \frac{\partial}{\partial \mathrm{b}_{jd}} \left[\log p({\mathrm{A}_j}^{-1}(\vec{X}_i - \vec{b}_j)) - \log \left| \det \mathrm{A}_j \right| \right] \tag{35}$$

An approximate method can also be applied for the adaptation of bias vectors instead of applying gradient. For approximate method, maximum likelihood estimate must satisfy the following condition:

$$\sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\Theta_j} \log p(\vec{X}_i|\hat{\xi}_j) = 0 \tag{36}$$

The bias term b_{jd} can be adapted as follows:

$$\nabla_{\mathrm{b}_{jd}} \mathcal{L}(\Theta, \mathcal{Z}, \mathcal{X}) = 0, \quad \Rightarrow \quad \sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\mathrm{b}_{jd}} \log p(\vec{X}_i|\xi_j) = 0 \tag{37}$$

By substituting Eq. (26) into Eq. (37), it is clear that gradient of the $\log p({\mathrm{A}_j}^{-1}(\vec{X}_i - \vec{\mathrm{b}}_j))$ must be zero as given in Eq. (38).

$$\nabla_{\mathrm{b}_{jd}} \log p({\mathrm{A}_j}^{-1}(\vec{X}_i - \vec{\mathrm{b}}_j)) = 0 \tag{38}$$

In adaptation of bias vector, if we assume that we have a large amount of data and prior probability distribution function of the source is differentiable and symmetric, then the $\log p({\mathrm{A}_j}^{-1}(\vec{X}_i - \vec{\mathrm{b}}_j))$ will be symmetric as well and the bias vector $\vec{\mathrm{b}}_j$ will be approximated by the weighted average of data samples as:

$$\vec{\mathrm{b}}_j = \frac{\sum_{i=1}^{N} \vec{X}_i \, p(j|\vec{X}_i)}{\sum_{i=1}^{N} p(j|\vec{X}_i)} \tag{39}$$

(c) *Shape Parameter Estimation* For estimation of parameters in ICA mixture model, unit variance and zero mean is assumed. For the purpose of estimation of shape parameter, same assumption is adopted and the problem will become estimation of shape parameter from data. The gradient ascent is used to estimate the shape parameter by maximizing log-likelihood:

$$\nabla_{\lambda_{jd}} \mathcal{L}(\Theta, \mathcal{Z}, \mathcal{X}) = \sum_{i=1}^{N} p(j|\vec{X}_i) \nabla_{\lambda_{jd}} \log p(\vec{X}_i|\xi_j) \tag{40}$$

The gradient ascent is used for adaptation, with gradient of the component density with respect to shape parameter vector λ_{jd} for each component of the mixture model.

$$\Delta\lambda_{jd} \propto p(j|\vec{X}_i) \frac{\partial}{\partial\lambda_{jd}} \log p(\vec{X}_i|\xi_j) \tag{41}$$

In adaptation of shape parameter λ_{jd}, gradient of component of the mixture model with respect to shape parameter is weighted by $p(j|\vec{X}_i)$. An estimate of shape parameter using gradient ascent is as follows:

$$\hat{\lambda}_{jd} = \lambda_{jd} + \alpha \left(p(j|\vec{X}_i) \frac{\partial}{\partial\lambda_{jd}} \log p(\vec{X}_i|\xi_j) \right) \tag{42}$$

The estimation of shape parameter in an ICA mixture model is discussed in [38] and the term $\frac{\partial}{\partial\lambda_{jd}} \log p(\vec{X}_i|\xi_j)$ is computed with the assumption of unit variance and zero mean as follows:

$$\begin{aligned} \frac{\partial}{\partial\lambda_{jd}} \log p(X_{id}|\xi_j) &= \frac{\partial}{\partial\lambda_{jd}} \log \left[\frac{f_{ggd}(X_{id}|\xi_j) \mathrm{H}(X_{id}|j)}{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j) d\mathrm{u}} \right] \\ &= h(X_{id}|\xi_j) - \frac{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j) h(\mathrm{u}|\xi_j) d\mathrm{u}}{\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j) d\mathrm{u}} \end{aligned} \tag{43}$$

where the term $h(X_{id}|\xi_j)$ is represented as:

$$\begin{aligned} h(X_{id}|\xi_j) &= \frac{\partial}{\partial\lambda_{jd}} \log f_{ggd}(X_{id}|\xi_j) \\ &= \left[\frac{1}{\lambda_{jd}} + \frac{3}{2\lambda_{jd}} \left[\Psi(1/\lambda_{jd}) - \Psi(3/\lambda_{jd}) \right] \right] - A(\lambda_{jd}) \left| X_{id} \right|^{\lambda_{jd}} \log \left| X_{id} \right| \\ &\quad - A(\lambda_{jd}) \left(\frac{1}{2} \log \frac{\Gamma(3/\lambda_{jd})}{\Gamma(1/\lambda_{jd})} + \frac{1}{2\lambda_{jd}} \left[\Psi(1/\lambda_{jd}) - 3\Psi(3/\lambda_{jd}) \right] \right) \left| X_{id} \right|^{\lambda_{jd}} \end{aligned} \tag{44}$$

The term $h(\mathrm{u}|\xi_j)$ also follows computation presented in Eq. (44). The term $\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j) h(\mathrm{u}|\xi_j) d\mathrm{u}$ can be approximated similar to Eq. (16).

$$\int_{\Omega_{S_j}} f_{ggd}(\mathrm{u}|\xi_j) h(\mathrm{u}|\xi_j) d\mathrm{u} \approx \frac{1}{M} \sum_{m=1}^{M} h(s_{j_{md}}|\xi_j) \mathrm{H}(s_{j_{md}}|j) \tag{45}$$

The estimation of shape parameter can be expressed as follows:

$$\hat{\lambda}_{jd} = \lambda_{jd} + \alpha \left[p(j|\vec{X}_i) \left\{ h(X_{id}|\xi_j) - \frac{\sum_{m=1}^{M} h(s_{j_{md}}|\xi_j) \mathrm{H}(s_{j_{md}}|j)}{\sum_{m=1}^{M} \mathrm{H}(s_{j_{md}}|j)} \right\} \right] \tag{46}$$

The complete learning procedure for BGGMM with ICA is given in Algorithm 1, where t_{min} is minimum threshold used to examine convergence criteria in each iteration.

Algorithm 1 Model Learning with BGGMM with ICA

Input:Dataset $\mathcal{X} = \{\vec{X}_1, \ldots, \vec{X}_N\}, t_{min}$.
Output: Θ.
{Initialization}: K-Means Algorithm. Set $\vec{\lambda}_j = 2$.
while relative change in log-likelihood $\geq t_{min}$ do
  {[**E Step**]}:
    for all $1 \leq j \leq K$ do
      Compute $p(j|\vec{X}_i)$ for $i = 1, \ldots, N$. using Eq. (10).
    end for
  {[**M step**]}:
    for all $1 \leq j \leq K$ do
      start ICA Algorithm
        Update the basis functions A_j using Eq. (31).
        Update the bias vector $\vec{b}_j$ using Eq. (39).
        Update shape parameter $\vec{\lambda}_j$ using Eq. (42).
      end ICA
      Update the mixing parameter p_j using Eq. (14).
      Update the mean $\vec{\mu}_j$ using Eq. (18).
      Update standard deviation $\vec{\sigma}_j$ using Eq. (21).
    end for
end while

3 Experiments and Results

The proposed algorithm is applied to BSS and unsupervised keyword spotting and experimental results are presented in the following subsections.

3.1 Blind Source Separation

3.1.1 Design of Experiments

In this subsection, experimental framework for BSS is described. It uses ICA mixture model for statistical learning as described in Sect. 2. In BSS, basis functions are estimated using ICA mixture model which is further applied to separate mixed signals. We have estimated basis functions 2×2, 3×3, 4×4 and 5×5 to compute 2, 3, 4 and 5 sources in separate experiments. In order to validate this BSS framework, TIMIT, TSP and NOIZEUS speech corpora are adopted during the experiments [23,30,36]. For BSS, only speech signal after linear mixing are observed. No prior information about basis functions is utilized during the source separation. BSS framework is evaluated using subjective and objective measures. Subjective analysis consists of speech signals before and after the source separation. Objective analysis consists of SDR, SIR, SAR and PESQ. Objective measures SDR, SIR and SAR are measured in dB and PESQ score lies in the range -0.5 to 4.5. Further details on objective measures can be found in [20,49,74]. This framework is also implemented using ICA in order to compare and examine the validity of statistical learning of ICA mixture model in BSS. ICA used in this work is implemented using Infomax [9].

Table 1 Objective measure for separation of 2 speech signals

Measure	TIMIT		TSP		NOIZEUS	
	ICA mix	ICA	ICA mix	ICA	ICA mix	ICA
SDR (dB)	61.57	57.42	60.28	55.47	51.86	45.38
SIR (dB)	62.29	55.89	61.15	57.91	47.95	43.53
SAR (dB)	292.75	276.75	295.81	279.19	289.48	280.69
PESQ	2.40	1.80	2.30	1.90	2.35	2.15

Table 2 Objective measure for separation of 3 speech signals

Measure	TIMIT		TSP		NOIZEUS	
	ICA mix	ICA	ICA mix	ICA	ICA mix	ICA
SDR (dB)	59.42	54.87	55.31	45.67	41.25	36.75
SIR (dB)	61.13	55.93	53.48	48.28	42.13	35.77
SAR (dB)	293.41	276.36	291.48	279.29	261.19	258.15
PESQ	2.35	1.65	2.40	1.80	2.20	1.90

3.1.2 Experimental Results

Blind source separation based on ICA mixture model is validated using TIMIT, TSP and NOIZEUS speech corpora. We have conducted 4 experiments to compute 2, 3, 4 and 5 speech sources from this BSS framework. For the recovery of 2, 3, 4 and 5 speech sources, we have taken linear mixture of 2, 3, 4 and 5 sources, respectively, from each database and performed blind source separation by employing BGGMM using ICA. Once the sources are recovered, objective analysis is performed on sources to examine quality of recovered speech signals and viability of ICA mixture model in BSS. Objective measures include SDR, SIR, SAR and PESQ analysis. SDR is a measure of distortion in output signal and it is defined as ratio between energy of clean signal and distortion and it is measured in dB. SIR is the ratio of target signal prower to the interference signal. It measures the amount of undesired interference still present after BSS and it is measured in dB. SAR measures the quality after source separation in terms of absence of artificial noise and measured in dB. PESQ is an objective assessment tool which correlates well with subjective listening scores [20,49,74]. The experiments are repeated 10 times with different linear speech mixtures of 2 and 3 sources from each database and average of objective measures is computed. In BSS for reconstruction of 4 and 5 sources, the experiments are repeated 10 times for TIMIT, 30 and 10 times for TSP and 7 and 6 times (due to the limitation of database) for NOIZEUS speech corpus, respectively and average of the objective measures is computed. For the case of TSP database with 4 sources, initially we repeated the experiment 10 times, but the effectiveness of our approach was not clear. We ran the same experiment for another set of 20 mixtures of 4 sources and averaged over all of 30 mixtures. We have performed same analysis using ICA in order to have a comparison of proposed BSS framework. The objective measures after the recovery of speech source signals are given in Tables 1, 2, 3 and 4. From the objective measures, it is observed that ICA mixture model outperforms the ICA in a relative setting of BSS for 2, 3, 4 and 5 sources for all databases. We do not have any permutation ambiguity while recovering 2 sources. In case of 3, 4 and 5 sources, permutation ambiguity is present

Table 3 Objective measure for separation of 4 speech signals

Measure	TIMIT		TSP		NOIZEUS	
	ICA mix	ICA	ICA mix	ICA	ICA mix	ICA
SDR (dB)	52.91	41.24	53.69	42.39	38.63	35.48
SIR (dB)	50.24	43.17	51.26	45.76	37.14	34.00
SAR (dB)	292.58	278.00	288.17	277.86	263.35	249.85
PESQ	2.10	2.00	2.05	1.85	2.20	1.85

Table 4 Objective measure for separation of 5 speech signals

Measure	TIMIT		TSP		NOIZEUS	
	ICA mix	ICA	ICA mix	ICA	ICA mix	ICA
SDR (dB)	51.87	46.44	52.92	40.55	49.75	38.04
SIR (dB)	52.65	47.16	49.21	39.49	50.71	40.63
SAR (dB)	291.29	277.32	285.03	276.55	271.19	260.12
PESQ	2.15	1.65	1.90	1.60	2.00	1.70

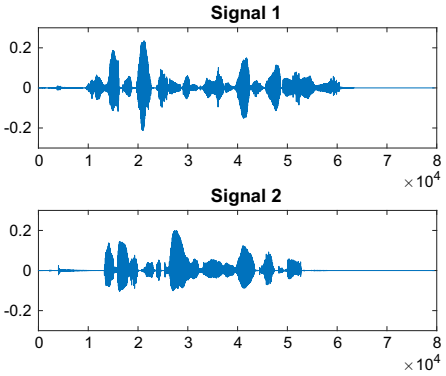
and it is higher while recovering higher number of sources. The speech signals before mixing, after mixing and after BSS are shown in Figs. 1 and 2, where we have plotted the signals in the same order as they are before mixing in order to present a clear comparison to the reader.

From the above experiments on BSS using BGGMM using ICA, it is observed that ICA mixture model performs better as compared to ICA. It is also observed that rate of this improvement becomes slower when we increase the number linear mixtures in source separation.

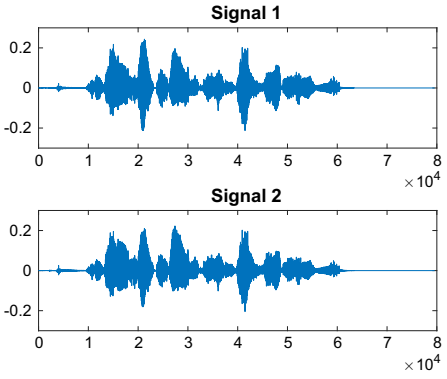
3.2 Blind Source Separation as Preprocessing to Keyword Spotting

In this subsection, proposed framework for BSS as pre-processing to unsupervised keyword spotting using an ICA mixture is presented. In real time applications, detection rate of speech recognition and keyword spotting is badly affected by mixing of speech signals with noise or other speech signals. It is also possible to intentionally mix the speech signal with noise or some other speech utterances to reduce or some times completely eliminate the chances of getting spotted by keyword spotting systems. In many security application of keyword spotting, it becomes critically important to use BSS as pre-processing when we are interested to spot particular keywords and we do not want to lose any piece of information.

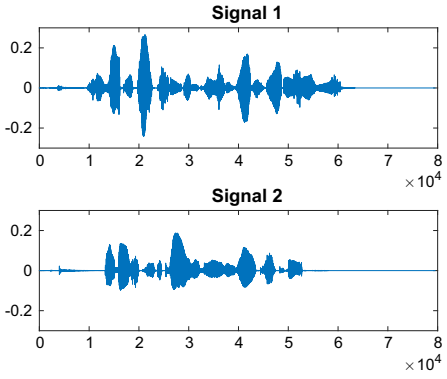
An unsupervised keyword spotting framework via segmental DTW on Gaussian posteriorgrams was presented in [80]. However, instead of using independently trained phonetic recognizer or GMM, an ICA mixture was proposed for training the model and generation of posteriorgram in [6]. The training process involves directly modeling the speech without any transcription information using the proposed ICA mixture model. The trained model was further used to decode the keyword examples and test utterances in posteriorgrams. Segmental DTW was used to compare the posteriorgrams between keyword examples and the test utterances. The distortion scores were ranked for the most reliable warping path to achieve keyword detection [25,80]. The detailed description on the keyword spotting is provided in [6,80]. In the experimental setup presented in [6], parameters of keyword spotting frame-



(a) Unmixed Signals



(b) Before Separation

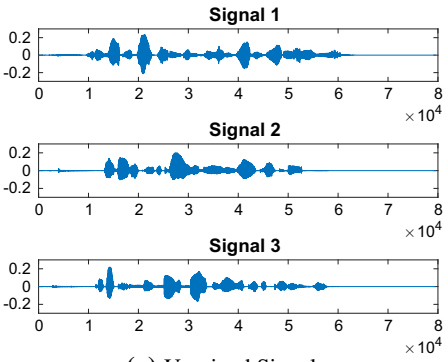


(c) After Separation

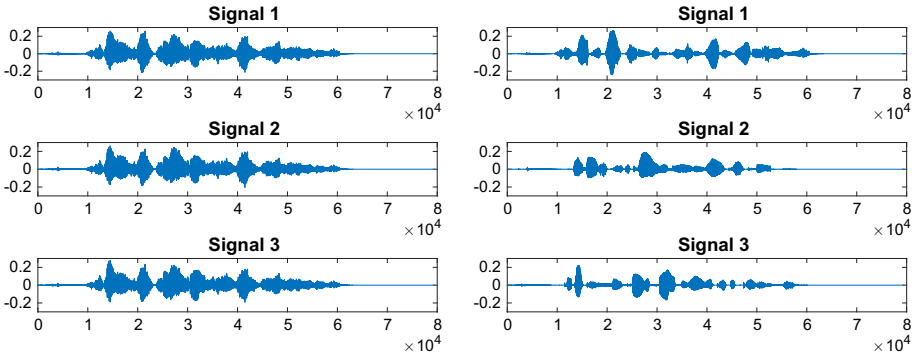
Fig. 1 Blind source separation with 2 signals

work were chosen exactly the same as given in [80], in order to have fair comparison of keyword detection results. ICA mixture model has demonstrated its viability and effectiveness in Keyword spotting framework based on detection rate presented in [6]. Experiments were performed on TIMIT speech corpus and a list of 10 keywords was selected to test the trained model for keyword spotting. In this framework same keyword spotting based on ICA mixture is adopted.

We have extended BSS framework presented in Sect. 3.1 and proposed as pre-processing for keyword spotting when the speech utterances with target keywords are mixed with noise or other speech utterances. The training phase of this proposed framework will remain same as presented in [6]. In order to examine the performance of keyword spotting framework, BSS is applied on test data to recover the speech signals. Once source separation is achieved through BSS via ICA mixture, the recovered signals can be applied to trained model for keyword detection. The proposed framework is shown in Fig. 3, which is inspired by [79]. Two types of problems occur in keyword spotting, when source mixing between speech utterances exist at initial stage during the testing. In the first case, if a speech utterance with a particular keyword is mixed with another speech utterance(s) and an overlap of a word exist in the second utterance(s) on same place as the particular keyword in first utterance. In



(a) Unmixed Signals



(b) Before Separation **(c)** After Separation

Fig. 2 Blind source separation with 3 signals

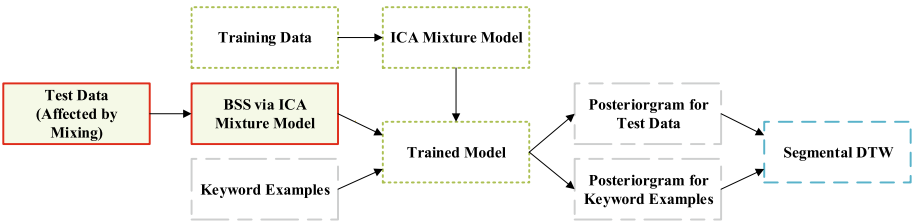


Fig. 3 Blind source separation as pre-processing to unsupervised keyword spotting via an ICA mixture model

this case, keyword will be mixed with the word of second utterance and it will more likely not detected during the keyword spotting. In second case, if a silent patch of speech exist in second utterance at the same place as keyword in the first speech utterance, it will be detected in the first speech utterance during keyword spotting. But it will also get detected in the second utterance which is a false alarm because keyword actually do not exist in the second speech utterance. These issues are addressed by proposing BSS as pre-processing to keyword spotting.

3.2.1 Design of Experiment

In this subsection, experimental framework and detection results for BSS based keyword spotting are presented. For keyword spotting, we have adopted the framework proposed in [6], and for the pre-processing stage, blind source separation framework presented in Sect. 3.1 is adopted. In both frameworks, ICA mixture is employed for statistical modeling and estimation of basis functions. In the training phase, speech data dedicated for training is used directly for statistical modeling without any transcription information. Once the model is trained, it can be used further to decode the keyword examples and test utterances into posteriorgrams. In this framework, it is assumed that test data is mixed with noise or other speech signals which requires the application of BSS before generation of posteriorgrams by employing the trained model. In order to perform pre-processing through BSS, we have created mixtures of 2, 3, 4 and 5 speech signals on test data. TIMIT speech corpus is employed during the modeling of keyword spotting framework and validation of the said framework is performed through the selected part of test data after being processed through BSS [23]. The speech signals processed through BSS are further applied to the trained model for generation of posteriorgrams. Segmental DTW is employed to compare the posteriorgrams for test utterances and keyword examples. Mel frequency cepstral coefficients (MFCCs) are used as features for in this framework.

3.2.2 Experimental Framework and Results

The BSS based keyword spotting framework is evaluated on TIMIT speech corpus. The TIMIT speech corpus is composed of 6300 speech utterances which contains 4620 speech utterances for training and 1680 speech utterances for testing. In this work, keyword spotting framework is modeled by all of the training data. For testing, speech utterances with target keywords and without target keywords were selected. The speech utterances with target keywords were mixed with the speech utterances without target keywords, for creating a mixture of 2, 3, 4, and 5 speech files. In these mixtures only one speech utterance have target keyword while rest of the speech utterances have no target keyword. Voice activity detection and feature extraction is applied directly before the modeling during training. For testing, feature extraction is applied after the test data is being processed through blind source separation. For feature extraction, each speech utterance is segmented into frames of 25 ms with a window shifting of 10 ms, where each frame is represented by 13 MFCCs. In order to initialize the parameters of ICA mixture during the training, K-Means is applied for mean, standard deviation and mixing weight whereas shape parameter is set to 2 for each component of mixture model. During the training for Keyword spotting, number of components of ICA mixture is set to be 50 as in [6,80]. The smoothing factor, segmental DTW adjustment window size and score weighing factor are chosen to be 0.00001, 6 and 0.5, respectively as in [6,80]. The keyword "Year" is uttered 177 times in the test part of dataset but in these experiments only 20 speech utterances with this keyword were selected, because rest of the keywords are uttered less than 20 times in the test data. For the testing, 10-keyword set presented in [6,80] is adopted and given in Table 5.

For the evaluation of keyword detection, three different evaluation matrices reported in [6,25,80] are examined, which are defined as: (1) the average precision for top 10 utterance hits termed as P@10, (2) the average precision for top N utterance hits termed as P@N, where N is equal to the number of occurrences of each keyword in the test data, (3) the average equal error rate (EER), where false acceptance rate is equal to false rejection rate. It is assumed that test data is affected by source mixing and it needs to be processed through BSS before

Table 5 TIMIT 10 keyword list used in [6,80]

age(3:10)	warm(10:8)	year(11:20)	money(19:17)	artists(7:7)
problem(22:9)	children(18:15)	surface(3:7)	development(9:8)	organizations(7:7)

applying to the trained model for generation of posteriorgrams. In order to validated the effectiveness of BSS as pre-processing, a new test data from the selected part of test data from TIMIT speech corpus is created. The purpose of this new dataset is to create the mixtures of 2×2, 3×3, 4×4 and 5×5 with speech utterances having target keywords and having no target keywords. In each mixture, one speech utterance have the target keyword while rest of them do not have target keyword. For example, in the case of keyword "age", all the 10 speech utterances with this keyword are taken and each utterance is mixed with another speech utterance with no target keyword for creating a mixture of 2×2. For mixture of 3×3, each speech utterance of target keyword is mixed with 2 more utterances having no target keyword. For the keyword "age", with mixtures of 2×2, we have 20 speech utterances in total (10 of them have target keyword and 10 have no target keyword), whereas with mixtures of 3×3, we have 30 speech utterances in total (10 of them have target keyword and 20 have no target keyword). All mixtures for the keywords given in Table 5 were created in the same fashion as discussed before. During the whole experiment, 100 speech utterances with no target keywords from the Table 5 and all the speech utterances with target keywords were selected and used according to the requirement for creating the mixture of speech data for each keyword. The next stage is to apply BSS and then adopt trained ICA mixture to generate posteriorgrams. BSS is performed by ICA mixture and same framework is adopted for BSS as discussed in Sect. 3.1. Table 6 indicates the performance of keyword detection before and after BSS, for different number of keyword examples based on P@N, P@N and EER.

For P@10 evaluation, 4 keywords from Table 5 are considered because only they have occurred more than 10 times both in the training and test part of dataset. For P@N and EER evaluations, with one keyword example experiment, all the keywords from the Table 5 were used. For P@N and EER evaluations, with 5 keyword examples experiment, 8 keywords were used because they have occurred more than 5 times in the training set. For P@N and EER evaluations, with 10 keyword examples experiment, only 5 keywords were used because they have occurred more than 10 times in the training set. The average precision for each keyword is calculated first and then mean of average precisions of all keywords for P@10 or P@N was computed. The EER for each keyword was computed based on false acceptance rate (FAR) and false rejection rate (FRR). The EER mentioned in Table 6 is the average of EER for all keywords used for that particular case [6,80]. Table 6 indicates considerable improvement in the evaluation matrices after being processed through BSS for 2×2 and 3×3 mixtures in comparison to the case when no BSS was applied. There is also improvement for 4×4 and 5×5 mixtures as compared to the case when no BSS was applied, but the trend of improvement is slow as compared to the 2×2 and 3×3 mixtures.

The results for average precisions (P@10 and P@N) are very close to each other, because utterance of available keywords in the test data is very close to 10 in most of the case. It is also important to note that only 4 keywords are present more than 10 times in both training and testing and hence P@10 was computed only for 4 keywords from the list given in Table 5. However for P@N, most of the keywords were used for computations, so it more effective for examining the viability of this framework. It is also observed that trend of improvement is higher when going from one keyword example to 5 keyword examples, whereas it is slow

Table 6 Evaluation matrix with BSS and without BSS

Mixture	Without BSS				After BSS		
	Examples	P@10 (%)	P@N (%)	EER (%)	P@10 (%)	P@N (%)	EER (%)
2 × 2	1	11.43	9.58	81.19	23.15	22.45	37.43
	5	13.76	12.25	77.55	46.86	43.89	25.38
	10	15.27	13.81	76.19	53.44	52.11	24.81
3 × 3	1	8.97	7.13	85.47	22.67	22.15	39.19
	5	10.14	9.58	81.44	44.63	41.74	26.88
	10	10.76	9.85	80.11	51.46	49.15	25.43
4 × 4	1	7.54	6.45	89.13	20.13	21.37	40.87
	5	8.10	6.93	87.46	40.92	38.49	29.15
	10	8.37	7.21	86.79	46.12	44.32	28.87
5 × 5	1	6.13	5.89	91.37	18.67	18.15	42.37
	5	6.48	6.27	88.49	38.19	36.56	31.13
	10	7.22	6.91	88.07	43.68	42.06	30.45

from 5 to 10 keyword examples. If we compare the results presented in this paper with the results presented in frameworks when no source mixing is considered, there is a lot of room for improvement. However comparison of keyword spotting with BSS and keyword spotting without BSS indicates the effectiveness of this framework in keyword spotting when speech signals are affected by mixing. It is also observed that the problem of false alarm due to mixing of sources is more severe in computing the detection rate for keyword spotting, which actually reduces the overall performance of keyword spotting. In many security applications, it is necessarily important to find the particular keywords because they are further used to detect particular speakers. If false alarm occur and even correct speaker is also detected, it will increase the number of possibilities to find the particular speaker. In the other case, when keyword is mixed with the words of other speech utterances and it is more likely not detected during the keyword spotting, it can increase the chances of completely losing a particular information. It is important when it is mixed intentionally to hide the particular information (keyword) which is critical to security. The experiments performed in this work only include the mixing of speech utterances. This framework needs to be extended for keyword spotting when speech utterances are mixed with noise. This experiment can be further extended with a larger vocabulary database by having more number of keyword examples.

4 Conclusion

In this paper, BGGMM with ICA is proposed as a model for statistical learning, which is further proposed in BSS and unsupervised keyword spotting to validate the effectiveness of algorithm. The proposed algorithm is a multivariate ICA mixture model which adopt EM and gradient ascent methods for parameter estimation. This algorithm handles the limitations associated with ICA which assumes that sources are independent from each other. In ICA mixture model, it is assumed that data come from a mixture model and it can be categorized into mutually exclusive classes with an ICA. The proposed model has demonstrated its success through presented applications. BSS adopts TIMIT, TSP and NOIZEUS speech corpora for

the validation and performance measures are computed using SDR, SIR,SAR and PESQ. From the experiments in BSS, it is observed that ICA mixture model performs better as compared to ICA. It is also observed that rate of this improvement becomes slower when we increase the number linear mixtures in source separation. BSS is further proposed as pre-processing to unsupervised keyword spotting, by employing ICA mixture model, when speech utterances having target keywords are affected by mixing of noise or other keywords. The experiments are performed by employing TIMIT speech corpus to train the ICA mixture for keyword spotting and then selecting the part of test data for creating a mixture of 2, 3, 4 and 5 speech signals to perform the blind source separation before the keyword spotting. The purpose of creating these mixtures of speech utterances with target keyword and with no target keyword is to validate the effectiveness of proposed framework. The keyword detection results are presented before and after the test data being processed through blind source separation. The keyword detection results based on average precision (P@10 and P@N), and EER validate the effectiveness of proposed framework when speech utterances with target keywords are affected by mixing. The experiments have shown significant improvement in the detection of keywords when mixed speech signals are processed through BSS via an ICA mixture. From the application of proposed algorithm in blind source separation and unsupervised keyword spotting, ICA mixture model has validated its effectiveness in statistical modeling.

Acknowledgements The completion of this research was made possible thanks to the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Alinaghi A, Jackson PJ, Liu Q, Wang W (2014) Joint mixing vector and binaural model based stereo source separation. IEEE/ACM Trans Audio Speech Lang Process 22(9):1434–1448. https://doi.org/10.1109/TASLP.2014.2320637
2. Allili M (2012) Wavelet modeling using finite mixtures of generalized gaussian distributions: application to texture discrimination and retrieval. IEEE Trans Image Process 21(4):1452–1464. https://doi.org/10.1109/TIP.2011.2170701
3. Allili M, Baaziz N, Mejri M (2014) Texture modeling using contourlets and finite mixtures of generalized Gaussian distributions and applications. IEEE Trans Multimed 16(3):772–784. https://doi.org/10.1109/TMM.2014.2298832
4. Allili MS, Bouguila N, Ziou D (2008) Finite general Gaussian mixture modeling and application to image and video foreground segmentation. J Electron Imaging 17(1):013,005–013,005
5. Ans B, Hérault J, Jutten C (1985) Adaptive neural architectures: detection of primitives. Proc COGNITIVA 85:593–597
6. Azam M, Bouguila N (2015) Unsupervised keyword spotting using bounded generalized Gaussian mixture model with ICA. In: 2015 IEEE GlobalSIP, 45: 1150–1154 . https://doi.org/10.1109/GlobalSIP.2015.7418378
7. Azam M, Bouguila N (2016) Speaker classification via supervised hierarchical clustering using ICA mixture model. Springer, Cham, pp 193–202. https://doi.org/10.1007/978-3-319-33618-3_20
8. Bae UM, Lee TW, Lee SY (2000) Blind signal separation in teleconferencing using ica mixture model. Electron Lett 36(7):680–682. https://doi.org/10.1049/el:20000459
9. Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7:1129–1159
10. Bishop CM (2006) Pattern recognition and machine learning (Information science and statistics). Springer, New York
11. Cardoso J (1997) Infomax and maximum likelihood for blind source separation. IEEE Signal Process Lett. https://doi.org/10.1109/97.566704
12. Choudrey RA, Roberts SJ (2003) Variational mixture of bayesian independent component analyzers. Neural Comput 15(1):213–252

13. Choy S, Tong C (2010) Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval. IEEE Trans Image Process 19(2):281–289. https://doi.org/10.1109/TIP.2009.2033400
14. Comon P (1992) Independent component analysis. In: International signal processing workshop on high-order statistics, Chamrousse, France, 10–12 July 1991, pp 111–120 (republished in J.L. Lacoume, ed., Hioher-Order Statistics, Elsevier, Amsterdam 1992, pp 29–38)
15. Comon P (1994) Independent component analysis, a new concept? Signal Process 36(3):287–314
16. Comon P, Jutten C (2010) Handbook of blind source separation: independent component analysis and applications, 1st edn. Academic Press, Cambridge
17. Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 28(4):357–366. https://doi.org/10.1109/TASSP.1980.1163420
18. Elguebaly T, Bouguila N (2014) Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection. Mach Vis Appl 25(5):1145–1162
19. Elguebaly T, Bouguila N (2015) Simultaneous high-dimensional clustering and feature selection using asymmetric Gaussian mixture models. Image Vis Comput 34:27–41
20. Emiya V, Vincent E, Harlander N, Hohmann V (2011) Subjective and objective quality assessment of audio source separation. IEEE Trans Audio Speech Lang Process 19(7):2046–2057. https://doi.org/10.1109/TASL.2011.2109381
21. Farag A, El-Baz A, Gimel'farb G (2006) Precise segmentation of multimodal images. IEEE Trans Image Process 15(4):952–968. https://doi.org/10.1109/TIP.2005.863949
22. Figueiredo MA, Jain AK (2002) Unsupervised learning of finite mixture models. IEEE Trans Pattern Anal Mach Intell 24(3):381–396
23. Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS, Dahlgren NL DARPA TIMIT acoustic phonetic continuous speech corpus CDROM. http://www.ldc.upenn.edu/Catalog/LDC93S1.html
24. Gu F, Zhang H, Wang W, Wang S (2017) An expectation-maximization algorithm for blind separation of noisy mixtures using Gaussian mixture model. Circuits Systems Signal Process 36(7):2697–2726. https://doi.org/10.1007/s00034-016-0424-2
25. Hazen T, Shen W, White C (2009) Query-by-example spoken term detection using phonetic posteriorgram templates. IEEE Workshop ASRU 2009:421–426. https://doi.org/10.1109/ASRU.2009.5372889
26. Hedelin P, Skoglund J (2000) Vector quantization based on gaussian mixture models. IEEE Trans Speech Audio Process 8(4):385–401. https://doi.org/10.1109/89.848220
27. Herault J, Jutten C (1986) Space or time adaptive signal processing by neural network models. In: Neural networks for computing, vol. 151, pp. 206–211. AIP Publishing, New York
28. Hérault J, Jutten C, Ans B (1985) Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In: 10 Colloque sur le traitement du signal et des images, FRA, 1985. GRETSI, Groupe dEtudes du Traitement du Signal et des Images
29. Hrault J, Ans B (1984) Circuits neuronaux synapses modifiables: dcodage de messages composites par apprentissage non supervis. C R Acad Sci 299:525–528
30. Hu Y, Loizou P (2007) Noizeus: A noisy speech corpus for evaluation of speech enhancement algorithms . http://ecs.utdallas.edu/loizou/speech/noizeus/. Online web resource
31. Huang X, Acero A, Hon H (2001) Spoken language processing: a guide to theory, algorithm, and system development, 1st edn. Prentice Hall PTR, Upper Saddle River
32. Hyvärinen A, Karhunen J, Oja E (2004) Independent component analysis, vol 46. Wiley, Hoboken
33. Jayashree P, Premkumar MJJ (2015) Machine learning in automatic speech recognition: a survey. IETE Tech Rev 0(0):1–12
34. Jutten C (1987) Calcul neuromimétique et traitement du signal: analyse en composantes indépendantes. Ph.D. thesis, Grenoble INPG
35. Jutten C, Herault J (1991) Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. Signal Process 24(1):1–10
36. Kabal P (2002) TSP speech database. Tech. rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada
37. Lee TW, Girolami M, Sejnowski TJ (1999) Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources
38. Lee TW, Lewicki MS (2000) The generalized Gaussian mixture model using ICA. In: International workshop on ICA, pp 239–244
39. Lee TW, Lewicki MS (2002) Unsupervised image classification, segmentation, and enhancement using ICA mixture models. IEEE Trans Image Process 11(3):270–279
40. Lee TW, Lewicki MS, Sejnowski TJ (1999) Unsupervised classification with non-Gaussian mixture models using ICA. In: Advances in neural information processing systems, pp 508–514

41. Lee TW, Lewicki MS, Sejnowski TJ (2000) ICA mixture models for unsupervised classification with non-Gaussian sources and automatic context switching in blind signal separation. In: IEEE transactions on pattern recognition and machine learning
42. Li W, Liao Q (2012) Keyword-specific normalization based keyword spotting for spontaneous speech. In: 8th international symposium on Chinese spoken language processing (ISCSLP), 2012, pp 233–237 https://doi.org/10.1109/ISCSLP.2012.6423490
43. Lindblom J, Samuelsson J (2003) Bounded support Gaussian mixture modeling of speech spectra. IEEE Trans Speech Audio Process 11(1):88–99. https://doi.org/10.1109/TSA.2002.805639
44. Liu C, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions. ECM ECME Stat Sinica 5(1):19–39
45. Liu G, Wu J, Zhou S (2013) Probabilistic classifiers with a generalized Gaussian scale mixture prior. Pattern Recognit 46(1):332–345
46. McGraw-Hill: Keyword spotting. (n.d.) mcgraw-hill dictionary of scientific & technical terms, 6e. (2003). http://encyclopedia2.thefreedictionary.com/keyword+spotting. Retrieved on March 31 2015
47. McLachlan G, Peel D (2004) Finite mixture models. Wiley, Hoboken
48. Mollah MNH, Minami M, Eguchi S (2006) Exploring latent structure of mixture ica models by the minimum β-divergence method. Neural Comput 18(1):166–190
49. Mowlaee P, Saeidi R, Christensen MG, Martin R (2012) Subjective and objective quality assessment of single-channel speech separation algorithms. In: 2012 IEEE ICASSP, pp 69–72
50. Myers C, Rabiner L (1981) A level building dynamic time warping algorithm for connected word recognition. IEEE Trans Acoust Speech Signal Process 29(2):284–297. https://doi.org/10.1109/TASSP.1981.1163527
51. Nguyen TM, Wu QJ, Zhang H (2014) Bounded generalized Gaussian mixture model. Pattern Recognit 47(9):3132
52. Palmer JA, Kreutz-delgado K, Makeig S (2006) An independent component analysis mixture model with adaptive source densities. Technical Report, UCSD
53. Park A, Glass J (2005) Towards unsupervised pattern discovery in speech. In: IEEE Workshop on automatic speech recognition and understanding, 2005, pp 53–58 https://doi.org/10.1109/ASRU.2005.1566529
54. Park A, Glass J (2006) Unsupervised word acquisition from speech using pattern discovery. In: IEEE proceedings of international conference on acoustics, speech and signal processing ICASSP, 2006, vol 1, pp. I–I . https://doi.org/10.1109/ICASSP.2006.1660044
55. Park A, Glass J (2008) Unsupervised pattern discovery in speech. IEEE Trans Audio Speech Lang Process 16(1):186–197. https://doi.org/10.1109/TASL.2007.909282
56. Peel D, McLachlan G (2000) Robust mixture modelling using the t distribution. Stat Comput 10(4):339–348
57. Peng T, Chen Y, Liu Z (2015) A time-frequency domain blind source separation method for underdetermined instantaneous mixtures. Circuits Syst Signal Process. https://doi.org/10.1007/s00034-015-0035-3
58. Persia LD, Milone D, Rufiner HL, Yanagida M (2008) Perceptual evaluation of blind source separation for robust speech recognition. Signal Process 88(10):2578–2583
59. Petersen KB, Winther O (2005) The EM algorithm in independent component analysis. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP '05), vol 5, pp v/169–v/172. https://doi.org/10.1109/ICASSP.2005.1416267
60. Price M, Glass J, Chandrakasan A (2015) A 6 mW, 5,000-word real-time speech recognizer using WFST models. IEEE J Solid-State Circuits 50(1):102–112. https://doi.org/10.1109/JSSC.2014.2367818
61. Rabiner L, Juang BH (1993) Fundamentals of speech recognition. Prentice-Hall Inc, Upper Saddle River
62. Ribeiro PB, Romero RAF, Oliveira PR, Schiabel H, Verosa LB (2013) Automatic segmentation of breast masses using enhanced ICA mixture model. Neurocomputing 120:61–71
63. Rohlicek J, Russell W, Roukos S, Gish H (1989) Continuous hidden markov modeling for speaker-independent word spotting. In: International conference on acoustics, speech, and signal processing, 1989. ICASSP-89, vol 1, pp 627–630. https://doi.org/10.1109/ICASSP.1989.266505
64. Rose R, Paul D (1990) A hidden Markov model based keyword recognition system. In: International conference on acoustics, speech, and signal processing, 1990. ICASSP-90., vol 1, pp 129–132. https://doi.org/10.1109/ICASSP.1990.115555
65. Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 26(1):43–49. https://doi.org/10.1109/TASSP.1978.1163055
66. Salazar A (2013) ICA and ICAMM methods. In: On statistical pattern recognition in independent component analysis mixture modelling, *Springer Theses*, vol 4. Springer, Berlin
67. Shah CA, Arora MK, Varshney PK (2004) Unsupervised classification of hyperspectral data: an ICA mixture model based approach. Int J Remote Sens 25(2):481–487

68. Shah CA, Varshney PK, Arora MK (2007) ICA mixture model algorithm for unsupervised classification of remote sensing imagery. Int J Remote Sens 28(8):1711–1731
69. Siu MH, Gish H, Chan A, Belfield W, Lowe S (2014) Unsupervised training of an HMM-based self-organizing unit recognizer with applications to topic classification and keyword discovery. Comput Speech Lang 28(1):210–223
70. Szoke I, Schwarz P, Burget L, Fapso M, Karafiat M, Cernocky J, Matejka P (2005) Comparison of keyword spotting approaches for informal continuous speech. In: In Proceedings, Interspeech
71. Takebayashi Y, Tsuboi H, Kanazawa H (1992) Keyword-spotting in noisy continuous speech using word pattern vector subabstraction and noise immunity learning. In: IEEE international conference on acoustics, speech, and signal processing, 1992. ICASSP-92, vol 2, pp. 85–88. https://doi.org/10.1109/ICASSP.1992.226114
72. Thiagarajan JJ, Ramamurthy KN, Spanias A (2013) Mixing matrix estimation using discriminative clustering for blind source separation. Digital Signal Process 23(1):9–18
73. Vincent E, Bertin N, Gribonval R, Bimbot F (2014) From blind to guided audio source separation: how models and side information can improve the separation of sound. IEEE Signal Process Mag 31(3):107–115. https://doi.org/10.1109/MSP.2013.2297440
74. Vincent E, Gribonval R, Fevotte C (2006) Performance measurement in blind audio source separation. IEEE Trans Audio Speech Lang Process 14(4):1462–1469. https://doi.org/10.1109/TSA.2005.858005
75. Wang H, Lee T, Leung CC, Ma B, Li H (2013) Unsupervised mining of acoustic subword units with segment-level Gaussian posteriorgrams. In: 14th annual conference of the international speech communication association INTERSPEECH 2013, Lyon, France, August 25–29, 2013, pp 2297–2301
76. Wei X, Yang Z (2012) The infinite student's t-factor mixture analyzer for Robust clustering and classification. Pattern Recognit 45(12):4346–4357
77. Wilcox L, Bush M (1992) Training and search algorithms for an interactive wordspotting system. In: IEEE international conference on acoustics, speech, and signal processing, 1992. ICASSP-92., 1992, vol 2, pp 97–100. https://doi.org/10.1109/ICASSP.1992.226111
78. Zhang Y (2009) Unsupervised spoken keyword spotting and learning of acoustically meaningful units. Master's thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science
79. Zhang Y (2013) Unsupervised speech processing with applications to query-by-exampley-example spoken term detection. Ph.D. thesis, MIT. Department of Electrical Engineering and Computer Science
80. Zhang Y, Glass J (2009) Unsupervised spoken keyword spotting via segmental DTW on Gaussian posteriorgrams. In: IEEE workshop on automatic speech recognition understanding, 2009. ASRU 2009. pp 398–403. https://doi.org/10.1109/ASRU.2009.5372931
81. Zhang Y, Glass J (2010) Towards multi-speaker unsupervised speech pattern discovery. In: IEEE international conference on acoustics speech and signal processing (ICASSP), 2010, pp 4366–4369 . https://doi.org/10.1109/ICASSP.2010.5495637
82. Zhang Y, Glass J (2011) An inner-product lower-bound estimate for dynamic time warping. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), 2011, pp 5660–5663 . https://doi.org/10.1109/ICASSP.2011.5947644