Aroudi et al. EURASIP Journal on Audio, Speech, and Music Processing (2015) 2015:35
DOI 10.1186/s13636-015-0078-1

EURASIP Journal on Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

Speech signal modeling using multivariate distributions

Ali Aroudi[1], Hadi Veisi[2*], Hossein Sameti[3] and Zahra Mafakheri[4]

Abstract

Using a proper distribution function for speech signal or for its representations is of crucial importance in statistical-based speech processing algorithms. Although the most commonly used probability density function (pdf) for speech signals is Gaussian, recent studies have shown the superiority of super-Gaussian pdfs. A large research effort has focused on the investigation of a univariate case of speech signal distribution; however, in this paper, we study the multivariate distributions of speech signal and its representations using the conventional distribution functions, e.g., multivariate Gaussian and multivariate Laplace, and the copula-based multivariate distributions as candidates. The copula-based technique is a powerful method in modeling non-Gaussian multivariate distributions with non-linear inter-dimensional dependency. The level of similarity between the candidate pdfs and the real speech pdf in different domains is evaluated using the energy goodness-of-fit test.

In our evaluations, the best-fitted distributions for speech signal vectors with different lengths in various domains are determined. A similar experiment is performed for different classes of English phonemes (fricatives, nasals, stops, vowels, and semivowel/glides). The evaluation results demonstrate that the multivariate distribution of speech signals in different domains is mostly super-Gaussian, except for Mel-frequency cepstral coefficient. Also, the results confirm that the distribution of the different phoneme classes is better statistically modeled by a mixture of Gaussian and Laplace pdfs. The copula-based distributions provide better statistical modeling of vectors representing discrete Fourier transform (DFT) amplitude of speech vectors with a length shorter than 500 ms.

Keywords: Multivariate distribution of speech signal, Copula-based multivariate distribution, Mel-frequency cepstral coefficient (MFCC), Discrete cosine transform (DCT), Discrete Fourier transform (DFT), Linear predictive coefficient (LPC), Goodness-of-fit (GOF) test

1 Introduction

Statistical-based speech processing algorithms have attracted wide interests during the last three decades in numerous applications, e.g., speech coding [1], speech recognition [2, 3], speech synthesis [4], and speech enhancement [5]. In all statistical-based speech processing algorithms, a probability density function (pdf) is assumed for the signal or its representation. Therefore, it is not surprising that proper selection of the pdf has been one of the challenges persistently addressed in this area [6–8].

Most of the statistical-based speech processing algorithms assume Gaussian probability distribution density for speech signals [2, 9–13]. The simplicity of the formulations and the semi-support of the central limit theorem are the main motivations for using Gaussian pdf [14]. Recently, a number of researchers have studied the distribution of speech signal more precisely using goodness-of-fit (GOF) test [15, 16] in time domain and transformed domains, e.g., discrete cosine transform (DCT) and discrete Fourier transform (DFT). In this regard, Gazor et al. [14], Martin [6], Shin et al. [7], Chen and Loizou [17], and Erkelens et al. [18] have shown that speech signals in various domains are modeled more accurately by super-Gaussian pdfs than by Gaussian pdf. Their evaluation results have demonstrated that the pdf of speech signals for time and DCT features are closer to Laplace [14], for jointly amplitude and phase of DFT features are closer to complex Laplace [17], for amplitude of DFT features are closer to Rayleigh [18], and for

* Correspondence: h.veisi@ut.ac.ir
[2]Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran
Full list of author information is available at the end of the article

real or imaginary parts of DFT features are individually closer to either Laplace or Gamma [6]. In addition, Shin et al. [7] have reported that generalized Gamma pdf models the distribution of real parts of DFT features more accurately compared to the Gaussian pdf. Table 1 summarizes the results of the published evaluations in this field.

All aforementioned publications have aimed to address the issue of modeling univariate pdf of speech signals for algorithms using univariate pdf. However, there are many statistical-based algorithms that take advantage of multivariate distribution of speech signals, and therefore, studying the multivariate distribution of speech to exploit a more proper pdf is a key issue for those speech processing algorithms too. There are typically several challenges in the studying and modeling of speech signals in the multivariate distribution case, e.g., the non-linear or linear inter-dimensional dependency, and the sparsity and complexity of the multidimensional space. These issues may have caused to mostly focus on the investigation of univariate distribution during the last two decades, and a small progress has been made in the multivariate distribution study of speech signal. The earlier studies on multivariate distribution of speech signal, performed by Brehm et al. [19] and LeBlancin et al. [20], suggested the multivariate Gaussian pdf for speech frames with a length of 5 ms. As the frame length and the process domain may vary the distribution [14], the multivariate Gaussian pdf may not be an appropriate choice for the algorithms using frame length other than 5 ms, e.g., 10 to 35 ms or exploiting process domain other than the time domain, e.g., DFT or DCT. In recent studies, Gazor et al. [14] and Jensen et al. [21] have used the moment test and have shown that Laplace multivariate distribution models speech signal or its representations are better than the Gaussian multivariate distribution. However, in these studies, the moment test as a GOF test was applied to each dimension individually and the possible contribution of inter-dimensional dependency to the multivariate distribution was not considered.

Table 1 Proposed super-Gaussian univariate distribution of speech signal in different domains

Domain	Fitted distribution
Time	Laplace [14]
Jointly amplitude and phase of DFT	Complex Laplace [17]
Amplitude of DFT	Rayleigh [18]
Real and imaginary parts of DFT	Laplace, Gamma [6], or generalized Gamma [7]
DCT	Laplace [14]

In this paper, we investigate multivariate distribution of speech signal in the time and transformed domains. We consider new plausible distribution candidates to tackle the multivariate distribution modeling challenges. Among the candidates, copula-based distributions are also proposed which are able to model the high-dimensional non-Gaussian distribution with non-linear inter-dimensional dependency [22]. The copula-based distributions have been popular over the last decade in the statistical fields, e.g., climate research, econometrics, risk management [22, 23], and finance [24, 25]. The other possible pdf candidates of speech including multivariate Gaussian, multivariate Laplace, the mixture of Gaussian, and the mixture of Laplace distributions are investigated in this paper too. We employ the goodness-of-fit test [15, 16, 26] to evaluate the degree of similarity between the candidate distribution and the real speech signal distribution. The GOF test is a tractable three-step approach to investigate distribution of data. In the first step, a number of candidates are assumed as the pdf of the real data. Next, an estimator, e.g., maximum likelihood (ML) is exploited to fit the candidates to the real data, and finally, the GOF test is performed to quantify the level of similarity between the fitted candidates and the real data. It is noted that although a wide number of GOF tests have been proposed, the most appropriate GOF test is the one that can highly cover underlying problem conditions, e.g., in our case study is high dimensionality of spaces. We briefly present a number of GOF tests, a summary of their strengths and deficiencies, and finally choose the one that has been reported as the most appropriate for high-dimensional space.

In general, speech processing algorithms using multivariate distribution exploit different feature types to process speech signals. For instance, traditional hidden Markov model (HMM)-based speech recognition and synthesis algorithms [3, 27] exploit Mel-frequency cepstral coefficients (MFCC); HMM-based speaker recognition [13] systems exploit either linear predictive coding (LPC) or MFCC; HMM-based speech enhancement algorithms use LPC, time, DCT, MFCC, or DFT [7, 9, 10]; and codebook-driven-based speech enhancement algorithms [28] employ LPC. However, all these algorithms assume the multivariate Gaussian pdf for extracted features of speech signals. As the feature type may influence the distribution [14], the multivariate distribution of the different feature types including DFT, DCT, time, LPC, and MFCC is studied in this paper. It is noted that a number of speech processing algorithms, e.g., proposed by Martin [6], Shin et al. [7], model the real and imaginary parts of DFT separately. Thus, we study the real and imaginary parts of DFT features separately. The whole study of multivariate distribution in this paper is concentrated on clean speech signals.

The remainder of this paper is structured as follows. In Section 2, the copula-based distributions are presented including their formulations and parameter estimation. In Section 3, the GOF tests are briefly reviewed and among them, the energy test is selected as the most appropriate one for the multivariate distribution study of high-dimensional space. Section 4 elaborates candidates' formulations, their parameter estimation, and an algorithm for exploring the best-fitted candidate. Section 5 presents the evaluation setup and experimental results. Finally, Section 6 concludes the work.

2 Copula-based distribution

A copula is defined as a multivariate probability distribution where the marginal probability distribution of each variable is uniform and is used to describe the dependency between random variables [22, 29–33]. As all the multivariate joint distributions can be written in terms of a copula and univariate marginal pdfs [29], copulas are used as a popular statistical tool for modeling multivariate distributions. In this regard, copulas allow to easily model the distribution of multivariate random variables by estimating only marginal pdfs and copulas. A copula-based distribution can capture important characteristics of a vector, e.g., the appropriate pdf for margins and the appropriate correlation structure with a possibly simple form.

The purpose of this section is to briefly review the basic definition of the copula and a number of the most commonly used estimation methods for fitting the copula to the real data.

2.1 Copula model

The mathematical basis of the copula was proposed by Sklar [29] and Hoeffding [33]. To define a copula model, let $\boldsymbol{x}$ be a d-dimensional random vector $\boldsymbol{x} = \{x_1, ..., x_j, ..., x_d\}$. Sklar [29] showed that the joint cumulative distribution function (CDF) of $\boldsymbol{x}$, $F_X(\boldsymbol{x})$, can be expressed as a function of marginal CDFs $u_j = F_{X_j}(x_j)$, $j = 1{:}d$, as shown in Eq. (1), where $C_X : [0,1]^d \rightarrow [0,1]$ denotes copula function. Based on Sklar's theorem [29], any arbitrary multivariate pdf $f_X(\boldsymbol{x})$ can be expressed as the product of two terms: the marginal pdf of dimensions $f_{X_j}(x_j)$, $j = 1{:}d$ and the copula density function $c_X(.)$ as shown in Eq. (2). The copula density function $c_X(.)$ can be derived from the copula function as given in Eq. (3). As the copula density function characterizes the inter-dimensional dependencies, it is also called correlation structure in literature [22].

$$F_X(\boldsymbol{x}) = C_X(u_1, ..., u_j, ..., u_d) \tag{1}$$

$$f_X(\mathrm{x}) = \left(\prod_{j=1}^{d} f_{X_j}(x_j)\right). c_X(u_1, ..., u_j, ..., u_d) \tag{2}$$

$$c_X(u_1, ..., u_j, ..., u_d) = \left(\prod_{j=1}^{d} \frac{\partial}{\partial u_j}\right). C_X(u_1, ..., u_j, ..., u_d) \tag{3}$$

The two most used parametric forms for $c_X(.)$ are elliptical and Archimedean [22, 30]. The Archimedean-based copulas are mostly used in the bivariate form and they are not practically usable for high-dimensional spaces due to its high-computational cost [22, 34]. In contrast to the Archimedean, the elliptical-based copulas, including Gaussian and Student-t copula, can be used for spaces with any number of dimensions [22]. We therefore briefly review the Gaussian and Student-t copulas in the following sections.

2.2 Gaussian copulas

Let us assume the marginal CDFs, u_j, are known. Each u_j can be transformed to a standard distributed random variable y_j, by using the inverse of univariate Gaussian CDF $F^{-1}_{N(0,1)}(.)$ as shown in the following equation,

$$y_j = F^{-1}_{N(0,1)}(u_j) = F^{-1}_{N(0,1)}\left(F_{X_j}(x_j)\right) \sim N(0,1). \tag{4}$$

As a consequence, the set $\boldsymbol{y} = \{y_1, ..., y_j, ..., y_d\}$ has a multivariate Gaussian distribution $N(0, \Sigma_{\mathrm{Copula}})$, where Σ_{Copula} denotes a symmetric, diagonal, positive definite matrix with unit diagonal elements. The joint CDF of $\boldsymbol{y}$ can therefore be expressed by Eq. (5), which can be interpreted as the copula function of $\boldsymbol{x}$ using terminology of Eq. (1). By using Eqs. (3) and (5), the copula density can be derived as shown in Eq. (6), where I denotes an identity matrix and Tr denotes transpose operator. For further details on this topic, see [22] and [31].

$$\begin{aligned} F_X(\boldsymbol{x}) &= C_X(u_1, ..., .u_j, .., u_d) \\ &= F_{\boldsymbol{N}(0,\Sigma_{\mathrm{Copula}})}(y_1, ..., y_j, ..., y_d) \end{aligned} \tag{5}$$

$$c_X(u_1, ..., u_d) = \frac{1}{\left|\Sigma_{\mathrm{Copula}}\right|^{0.5}} \exp\left(-\frac{1}{2}\, y\left(\Sigma^{-1}_{\mathrm{Copula}} - I\right) y^{Tr}\right) \tag{6}$$

2.3 Student-t copulas

The Student-t copula is defined analogously to the Gaussian copula; however, the transformed variables y_j are obtained using univariate Student-t CDF inverse $F^{-1}_{t(\nu)}(u_j)$, where $t(\nu)$ is a univariate Student-t distribution with ν degrees of freedom. Consequently, $\boldsymbol{y} = \{y_1, ..., y_j, ..., y_d\}$ follows a multivariate Student-t distribution $\boldsymbol{t}(\nu, \Sigma_{\mathrm{Copula}})$ where Σ_{Copula} denotes a positive definite matrix with unit diagonal elements. Similar to the Gaussian copula function, the Student-t copula function, i.e., joint CDF of $\boldsymbol{y}$, is shown by Eq. (7). By using Eqs. (3) and (7),

the Student-t copula density then is computed as shown in Eq. (8). For further details, see [22] and [31].

$$C_{\boldsymbol{X}}(u_1,...,u_d) = F_{\boldsymbol{t}\left(\nu,\Sigma_{\text{Copula}}\right)}\left(F^{-1}_{t(\nu)}(u_1),...,F^{-1}_{t(\nu)}(u_d)\right) \tag{7}$$

$$\begin{aligned} c_{\boldsymbol{X}}(u_1,...,u_d) &= \left|\Sigma_{\text{Copula}}\right|^{-0.5}\frac{\Gamma\left(\frac{\nu+d}{2}\right)\left[\Gamma\left(\frac{\nu}{2}\right)\right]^d}{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^d\Gamma\left(\frac{\nu}{2}\right)} \\ &\times\frac{\left(1+\frac{1}{\nu}y\Sigma^{-1}_{\text{Copula}}y^t\right)^{-\frac{\nu+d}{2}}}{\prod_{j=1}^{d}\left(1+\frac{\mathrm{Y}^2}{\nu}\right)^{-\frac{\nu+1}{2}}} \end{aligned} \tag{8}$$

2.4 Fit a copula model

There are several estimation methods proposed for copula parameters [22, 35–37]. Among them, maximum likelihood (ML), inference functions for margins (IFM), and canonical maximum likelihood (CML) techniques are used more frequently than others [22]. Generally, these estimators maximize the log-likelihood function of Eq. (2) with respect to $\Theta = \{\theta, \alpha\}$ in different ways, where $\boldsymbol{\alpha}$ denotes a set of parameters concerning the copula density function, e.g., Σ_{Copula}, and $\boldsymbol{\theta} = \{\theta_1, ..., \theta_j, ..., \theta_d\}$ denotes a set of parameters concerning the marginal pdfs, $f_{\mathrm{X}_j}(x_j)$. The log-likelihood function is derived as shown in Eq. (9) [31] where $\boldsymbol{x}_{n=1}^{N} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^n, ..., \boldsymbol{x}^N\}$ denotes a set of N observation vectors of real data and x_j^n represents jth component of vector $\boldsymbol{x}^n$.

$$\begin{aligned} \mathrm{l}(\boldsymbol{\Theta}) = &\sum_{n=1}^{N}\sum_{j=1}^{d}\ln f_{\mathrm{X}_j}\left(x_j^n;\theta_j\right) \\ &+\sum_{n=1}^{N}\ln c\left[F_{\mathrm{X}_1}(x_1^n;\theta_1),...,F_{\mathrm{X}_j}\left(x_j^n;\theta_j\right),...,F_{\mathrm{X}_d}(x_d^n;\theta_d);\boldsymbol{\alpha}\right] \end{aligned} \tag{9}$$

The ML approach estimates the parameters of marginal pdfs and copula density function jointly using numerical optimization [22]. This is the only way to estimate all the parameters consistently [22].

The IFM method is the most used method. It estimates by maximizing Eq. (9) in two steps. First, the parameters of the margins $\boldsymbol{\theta}$ are individually estimated as shown in Eq. (10). It is noted that the type of marginal distributions, e.g., Laplace or Gamma, are assumed to be determined in advance. The copula parameters $\boldsymbol{\alpha}$ are then derived as shown in Eq. (11) by using the estimated $\hat{\boldsymbol{\theta}}$.

$$\hat{\theta}_j = \arg\max\sum_{n=1}^{N}\ln f_{X_j}\left(x_j^n;\theta_j\right) \tag{10}$$

$$\hat{\boldsymbol{\alpha}} = \arg\max\left[\sum_{n=1}^{N}\ln c\left(F_{\mathrm{X}_1}\left(x_1^n;\hat{\theta}_1\right),...,F_{\mathrm{X}_j}\left(x_j^n;\hat{\theta}_j\right),...,F_{\mathrm{X}_d}\left(x_d^n;\hat{\theta}_d\right);\boldsymbol{\alpha}\right)\right] \tag{11}$$

The CML method first empirically estimates $f_{\mathrm{X}_j}\left(x_j^n\right)$ and $F_{\mathrm{X}_j}\left(x_j^n\right)$, denoted as $\hat{f}_{\mathrm{X}_j}\left(x_j^n\right)$ and $\hat{F}_{\mathrm{X}_j}\left(x_j^n\right)$, using non-parametric methods, e.g., kernel smoothing density estimator [38]. Thus, it is not needed to determine the type of marginal distributions in advance, in contrast to the IFM method. The parameters of copula are then estimated using Eq. (12).

$$\hat{\boldsymbol{\alpha}} = \arg\max\left[\sum_{n=1}^{N}\ln c\left(\hat{F}_{\mathrm{X}_1}(x_1^n),...,\hat{F}_{\mathrm{X}_j}\left(x_j^n\right),...,\hat{F}_{\mathrm{X}_d}(x_d^n);\boldsymbol{\alpha}\right)\right] \tag{12}$$

In order to estimate the parameters of the copula-based distribution using one of the discussed methods, the following issues should be considered:

- The ML method is used for spaces with a small number of dimensions due to the numerical complexity issue.
- The IFM method ends up a sub-optimal solution for parameter estimation since the log-likelihood function is maximized in two individual steps [31].
- The IFM and CML methods result in closed-form formulas only for Gaussian copula case [31].
- When one of the values of off-diagonal components of the covariance matrix ΣCopula of either the Student-t or Gaussian copulas takes 1 or −1, the estimation procedure of the copula parameters using CML method may fail [39]. It is due to Cholesky decomposition performed in CML methods.

In this paper, the parameters of copula-based distributions are estimated using IFM method and Gaussian copula density function that result in closed-form formulas and avoid failing in the estimation procedure. To estimate $\boldsymbol{\alpha} = \Sigma_{\text{Copula}}$ of Gaussian given a vector sequence $\boldsymbol{y}_{n=1}^{N} = \{\boldsymbol{y}^1, ..., \boldsymbol{y}^n, ..., \boldsymbol{y}^N\}$, the Eq. (6) is plugged into Eq. (11) resulting in Eq. (13). Regarding the estimation of $\boldsymbol{\theta}$, as the type of marginal pdfs of copula density function should be given in advance, it will be discussed in Section 4.

$$\hat{\Sigma}_{\text{Copula}} = \frac{1}{N}\sum_{n=1}^{N} (\boldsymbol{y}^n)^{Tr}\boldsymbol{y}^n \tag{13}$$

3 Goodness-of-fit test

A wide number of goodness-of-fit (GOF) tests depending on underlying conditions of the case study have been proposed. In our study, high dimensionality and possible non-linear inter-dimensional dependency are the most crucial issues. Although various GOF tests are proposed for one-dimensional space, only some of them are extendable for high-dimensional space. As the number of space dimensions increases, the tests become inefficient [26]. For instance, the extension of χ^2 test [15] to higher dimensions suffers from the curse of dimensionality [40] caused by the space sparsity unless the sample sizes are large enough.

There are several GOF tests particularly proposed for the multivariate case, e.g., the nearest neighbor test which exploits the nearest neighbors [41], Mardia test which exploits the skewness and kurtosis [42] and Freidman–Rafsky test which exploits the minimum spanning tree [43]. In this regard, the energy test has also recently been proposed by Zach and Aslan [44]. The performance superiority of the energy test has been demonstrated compared to Mardia, nearest neighbor, χ^2, and Friedman–Rafsky tests. Accordingly, the energy test is selected as a more appropriate GOF for underlying conditions of the study in this paper to evaluate candidates. In the following, the energy test is discussed.

3.1 Energy test

Given a candidate pdf $f_{X_0}(\boldsymbol{x})$ and a sequence of observation vectors $\boldsymbol{x}_{n=1}^N = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^i, ..., \boldsymbol{x}^N\}$ which follow an unknown pdf $f_X(\boldsymbol{x})$, the energy statistic for the hypothesis H_0: $fx(\boldsymbol{x}) = f_{X0}(\boldsymbol{x})$ against H_1: $fx(\boldsymbol{x}) \neq f_{X0}(\boldsymbol{x})$ is computed by

$$\begin{aligned}\phi_{NM} = &\frac{1}{N(N-1)}\sum_{t>i} R\left(\left|\boldsymbol{x}^i - \boldsymbol{x}^t\right|\right) \\ &+ \frac{1}{M(M-1)}\sum_{j>n} R\left(\left|q^n - q^j\right|\right) \\ &- \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} R\left(\left|\boldsymbol{x}^i - \boldsymbol{q}^j\right|\right)\end{aligned} \tag{14}$$

where $\boldsymbol{q}_{j=1}^M = \{\boldsymbol{q}^1, ..., \boldsymbol{q}^j, ..., \boldsymbol{q}^M\}$ denotes a sequence of M observation vectors following $f_{X_0}(\boldsymbol{x})$ and generated by Monte–Carlo simulation [45] and R denotes a continuous, monotonically decreasing function, i.e., $R(r) = -\ln(r)$. In the limit $N \to \infty$ and $M \to \infty$, the statistic ϕ_{NM} approaches minimized value, near zero, if $\boldsymbol{x}_{i=1}^N$ and $\boldsymbol{q}_{j=1}^M$ are from the same distribution [44]. For further details, see the Appendix.

The required steps for performing the energy test are summarized:

1. The real dataset is segmented, resulting in N vectors each of length d, $\boldsymbol{x}_{i=1}^N = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^i, ..., \boldsymbol{x}^N\}$. Depending on the process domain, $\boldsymbol{x}^i$ represents the segmented real data in that process domain, e.g., time, DFT, and DCT.
2. A possible candidate pdf $f_{X_0}(\boldsymbol{x})$, e.g., either copula-based or conventional distributions, is hypothesized and fitted to the real data vectors $\boldsymbol{x}_{i=1}^N$.
3. The number M of simulated data vectors following fitted pdf $f_{X_0}(\boldsymbol{x})$ is generated using Monte–Carlo.
4. The energy test statistic is computed using Eq. (14) to determine the level of similarity between the distributions of real data vectors $\boldsymbol{x}_{i=1}^N$ and simulated data vectors $\boldsymbol{q}_{j=1}^M$.

4 Multivariate distribution candidates

To study the multivariate distribution of speech features, two classes of pdfs are considered as the candidates in this paper: (1) copula-based distributions and (2) conventional distributions. The first pdf class includes five distributions:

1. Copula-based Laplace distribution (CLD)
2. Copula-based Laplace distribution with mutually independent dimensions (CLID), i.e., $c_X(.) = 1$
3. Copula-based generalized extreme value distribution (CGevD)
4. Copula-based Rayleigh distribution (CRD)
5. Copula-based Gamma distribution (CGD).

As formerly mentioned, the IFM method is used to fit the copula-based distributions to real data. In this regard, marginal distributions of each candidate should be first determined. The following univariate pdfs are used as the marginal distributions for each candidate: univariate Laplace pdf as shown in Eq. (15) [14] for CLD and CLID, univariate generalized extreme value pdf as shown in Eq. (16) [46] for CGevD, univariate Rayleigh pdf as shown in Eq. (17) for CRD, and univariate Gamma pdf as shown in Eq. (18) [21] for CGD. In these equations, μ_j and α_j denote the location parameters, b_j, ξ_j, and σ_j denote the scale parameters, and k_j and $\gamma_j$represent the shape parameters. As a consequence, CLD is a copula-based distribution with marginal distributions of Laplace and its parameters $\boldsymbol{\Theta}_{\text{CLD}} = \{\boldsymbol{\theta}_{\text{Laplace}}, \boldsymbol{\alpha}_{\text{Copula}}\}$ are estimated using the IFM method given an observation set of vectors $\boldsymbol{x}_{n=1}^N = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^i, ..., \boldsymbol{x}^N\}$, where $\boldsymbol{\theta}_{\text{Laplace}}$ denotes a parameter set $\{b_j, \mu_j\}$ concerning marginal distribution

$f_{X_j,\text{Laplace}}(x_j)$, $j = 1{:}d$ and is estimated using Eqs. (19) and (20). The parameter of Gaussian copula density $\boldsymbol{\alpha}_{\text{Copula}}$ is estimated using Eq. (13). Similarly, $\boldsymbol{\theta}_{\text{Rayleigh}} = \{\nu_j\}$ is estimated using Eq. (21) and $\boldsymbol{\theta}_{\text{Gamma}} = \{\gamma_j, \xi_j\}$ is estimated using Eqs. (22) and (23) [47], where $s = \ln\left(\frac{1}{N}\sum_{n=1}^{N} x_j^n\right) - \frac{1}{N}\cdot\sum_{n=1}^{N}\ln\left(x_j^n\right)$. It is noted that CGD and CRD both model one-side distributions and are therefore proposed as candidates for modeling one-side distributed random vector, e.g., the amplitude of DFT feature. For estimation of CGevD parameters $\boldsymbol{\theta}_{\text{GEV}} = \{\sigma_j, \alpha_j, k_j\}$ using the IFM, as it results in no closed-form solution [48], the MATLAB function "fminsearch", which employs numerical method of Lagarias [49], is employed.

$$f_{X_j,\text{Laplace}}(x_j) = \frac{1}{2b_j}.exp\left(-\frac{\left|x_j-\mu_j\right|}{b_j}\right) \tag{15}$$

$$f_{X_j,\text{GEV}}(x_j) = \left(\frac{1}{\sigma_j}\right).exp\left(-\left(1+k_j\frac{x_j-\alpha_j}{\sigma_j}\right)^{-\frac{1}{k_j}}\right).\left(1+k_j\frac{\left(x_j-\mu_j\right)}{\sigma_j}\right)^{-1-\frac{1}{k_j}} \tag{16}$$

$$f_{X_j,\text{Rayleigh}}(x_j) = \frac{x}{\nu_j^2}.exp\left(-\frac{x^2}{2\nu_j^2}\right) \tag{17}$$

$$f_{X_j,\text{Gamma}}(x_j) = \frac{(x_j)^{\gamma_j-1}}{(\xi_j)^{\gamma_j}\Gamma\left(\gamma_j\right)}.exp\left(\frac{-x_j}{\xi_j}\right) \tag{18}$$

$$\hat{\mu}_j = \text{median}\left(x_j^n\right) \tag{19}$$

$$\hat{b}_j = \frac{1}{N}\cdot\sum_{n=1}^{N}\left|x_j^n-\mu_j\right| \tag{20}$$

$$\hat{\nu}_j \approx \sqrt{\frac{1}{2N}\cdot\sum_{n=1}^{N}\left(x_j^n\right)^2} \tag{21}$$

$$\hat{\gamma}_j \approx \frac{3-s+\sqrt{(s-3)^2+24s}}{12s}\sqrt{\frac{1}{2N}\cdot\sum_{n=1}^{N}\left(x_j^n\right)^2} \tag{22}$$

$$\hat{\xi}_j = \frac{1}{\gamma_j N}\cdot\sum_{n=1}^{N} x_j^n \tag{23}$$

The following steps summarize the estimation procedure of Gaussian copula.

1. A marginal distribution, e.g., one of Eqs. (15)–(18), is accounted for $fX_j(x_j)$, $j = 1{:}d$ and parameter set is accordingly estimated.
2. The corresponding CDF of $f_{X_j}(x_j)$ is derived to compute $u_j = F_{X_j}(x_j)$, $j = 1{:}d$ and $y_j = F^{-1}_{N(0,1)}(u_j)$ as shown in Eq. (4).
3. The parameter of Gaussian copula density $\alpha_{\text{Copula}} = \Sigma_{\text{Copula}}$ is estimated using Eq. (13).

The second class of candidate distributions, i.e., conventional distributions, includes five conventional pdfs: multivariate Laplace (MLD), multivariate Gaussian (MGD), and three multivariate Gaussian–Laplace mixtures (MGLD). The MGD is considered as shown in Eq. (24) where the full covariance matrix Σ and mean vector μ are estimated using the maximum likelihood method [16]. Regarding MLD, as the ultimate purpose is to compute the energy test statistic using simulated vectors following MLD, the required vectors $\boldsymbol{s}^i_{\text{Laplace}}$ are generated using Eq. (25) [50] where $\boldsymbol{x}^i_{\text{Gaussian}}$ and w_i denote simulated vector following MGD of Eq. (24) and simulated sample following univariate exponential distribution of Eq. (26),

Table 2 Multivariate distribution candidates considered for experimental setup

PDF class	Candidates	Description
Copula-based PDF	CLD	Copula-based distribution with marginal Laplace distribution.
	CLID	Copula-based distribution with mutually independent marginal Laplace distribution.
	CGevD	Copula-based distribution with marginal GEV distribution.
	CRD	Copula-based distribution with marginal Rayleigh distribution.
	CGD	Copula-based distribution with marginal Gamma distribution.
Conventional PDF	MGD	Multivariate Gaussian distribution.
	MLD	Multivariate Laplace distribution.
	MGLD, $p = 0.25$	Multivariate Gaussian–Laplace distribution.
	MGLD, $p = 0.50$	Multivariate Gaussian–Laplace distribution.
	MGLD, $p = 0.75$	Multivariate Gaussian–Laplace distribution.

Table 3 The data set of the second setup of evaluations

File	Phoneme class	# of phonemes	Duration (s)	# of frames for each phoneme class (*N*)			
				20 ms	30 ms	100 ms	500 ms
1	Semivowel/glide	460	29.77	1489	993	298	60
2	Vowel	1240	125.39	6270	4180	1254	251
3	Nasal	298	18.57	929	620	186	38
4	Fricative	482	46.16	2309	1539	462	93
5	Stop	1109	56.13	2870	1872	562	113

respectively. For MGLD observations, the simulated vectors are generated using Eq. (27) where $\boldsymbol{s}^i_{\text{Laplace}}$ and $\boldsymbol{x}^i_{\text{Gaussian}}$ denote the simulated vectors following distributions of Eqs. (25) and (24), respectively. In this equation, variable p shows the amount of contribution of Laplace distribution compared with Gaussian distribution in generating $\boldsymbol{z}^i$. Three multivariate Gaussian–Laplace mixtures are considered with corresponding values of $p \in \{0.25,\ 0.50,\ 0.75\}$, denoted as MGLD with $p = 0.25$, MGLD with $p = 0.50$, and MGLD with $p = 0.75$, for the experimental evaluations.

$$f_{\boldsymbol{X},\,\text{Gaussian}}(\boldsymbol{x}) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}}.exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)\Sigma^{-1}(\boldsymbol{x}-\mu)^t\right) \quad (24)$$

$$\boldsymbol{s}^i_{\text{Laplace}} = \sqrt{w_i}\boldsymbol{x}^i_{\text{Gaussian}} \quad (25)$$

$$f_{\text{Exponential}}(w_i) = \frac{1}{\lambda}\exp\left(\frac{-w_i}{\lambda}\right) \quad (26)$$

$$\boldsymbol{z}^i = p\boldsymbol{s}^i_{\text{Laplace}} + (1-p)\boldsymbol{x}^i_{\text{Gaussian}} \quad (27)$$

Consequently, eight candidates are considered for multivariate distribution study of speech features, except for amplitude of DFT feature. For the amplitude of DFT feature, two additional candidates CGD and CRD resulting in total ten candidates are considered. Table 2 summarizes the candidates.

5 Evaluation results

In this section, the experimental evaluation results of the multivariate speech distribution study are presented. To perform the evaluations, 100 sentences, uttered by 11 male and female native English speakers (with New York City dialect region), with sampling rate of 16 kHz from TIMIT database [51] were randomly selected (see Additional file 1). Two experimental setups were considered for evaluations. For the first experimental setup, all speech information of 100 sentences was exploited for computing the statistic of the energy test. For the second experimental setup, phoneme-based evaluations were performed, i.e., five classes of English phonemes were used (fricatives, nasals, stops, vowels, and semivowel/glides). For each phoneme class, the relevant information was first extracted from the 100 sentences used in the first experimental setup and then concatenated to produce one file. As a result, five output files for five phoneme classes were produced. Table 3 shows the content and the number of extracted phonemes of each of the five files. The evaluation results of the second experimental setup benefit the statistical-based speech recognition and synthesis algorithms statistically modeling phoneme classes. As non-speech information (silence interval) may influence the distribution [14], it was removed for both experimental setups. The total duration of the first setup data after excluding silence intervals was 337.84 s. For the second setup, the duration of each file is represented in Table 3.

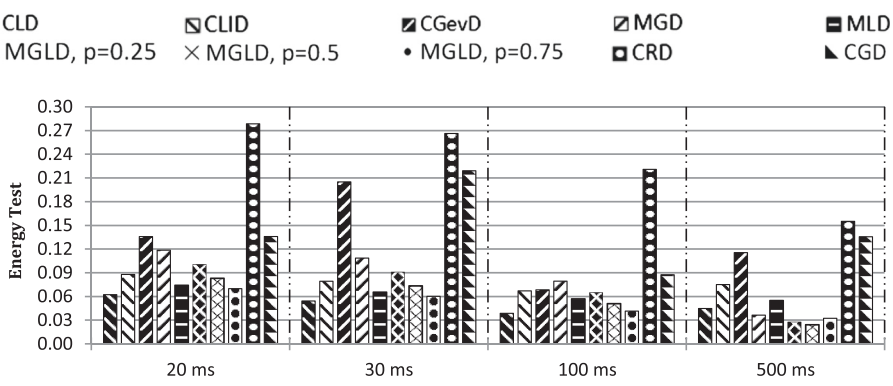

Fig. 1 Values of the energy test for different multivariate distributions (candidates) resulted from ADFT features with frame lengths of 20, 30, 100, and 500 ms

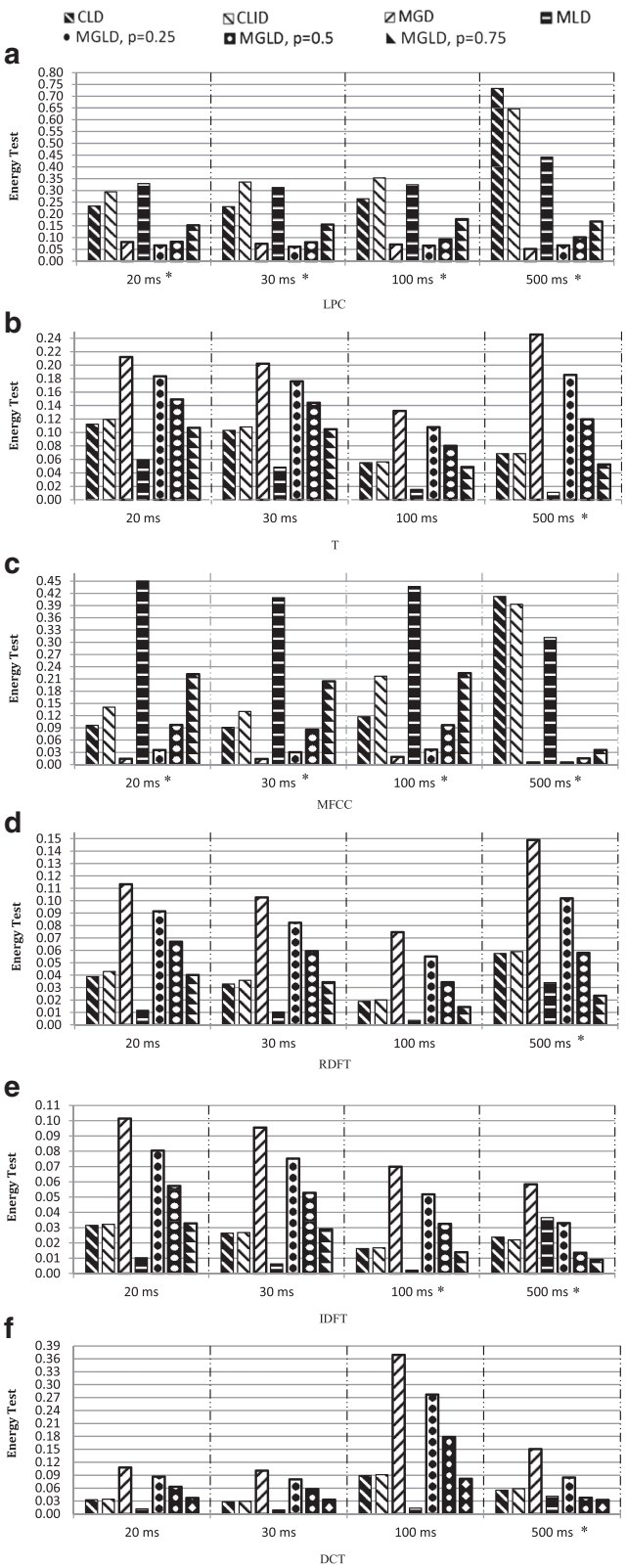

Fig. 2 Values of the energy test for different multivariate distributions (candidates) resulted from **a** LPC, **b** T, **c** MFCC, **d** RDFT, **e** IDFT, and **f** DCT features with frame lengths of 20, 30, 100, and 500 ms

Table 4 Best-fitted multivariate distribution in the sense of energy test for different features and frame lengths of speech signals

Feature (domain)	Frame length			
	20 ms	30 ms	100 ms	500 ms
ADFT	CLD	CLD	CLD	MGLD, $p = 0.50$
LPC	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGD
T	MLD	MLD	MLD	MLD
MFCC	MGD	MGD	MGD	MGD
RDFT	MLD	MLD	MLD	MGLD, $p = 0.75$
IDFT	MLD	MLD	MLD	MGLD, $p = 0.75$
DCT	MLD	MLD	MLD	MGLD, $p = 0.25$

Prior to performing evaluation, the datasets of both setups were segmented in frames with lengths of 20, 30, 100, and 500 ms. For the first setup, the number of frames, N, corresponding to the segment lengths of 20, 30, 100, and 500 ms were 13936, 9291, 2788, and 558, respectively. For the second setup, the number of frames, N, corresponding to the segment lengths for each phoneme class is shown in Table 3.

The experimental evaluations were performed for time features (T), amplitude of DFT (ADFT), real parts of DFT (RDFT), imaginary parts of DFT (IDFT), DCT, LPC, and MFCC features. Regarding the LPC and MFCC, 10 and 12 coefficients were extracted from frames, respectively. The MFCC vectors were extracted from 23 Mel-frequency filter banks. To set up the energy test, the value of M was taken equal to N. All the reported energy test values were computed at a significant level of 0.01 using a bootstrap method [44, 52].

Figure 1 represents the experimental results of the energy test for ADFT features, concerning the first setup. Fig. 2 illustrates the experimental results of the energy test for other features including LPC, T, MFCC, RDFT, IDFT, and DCT features. Regarding Fig. 2, as the energy test values of some cases were much lower in comparison with the others, they were scaled up ten times to be illustrated better and punctuated by * on the right side of frame length of horizontal axis, e.g., 500 ms *. Moreover, as the energy test values of CGevD for all features were far greater than the others, they were schematically removed from Fig. 2 to have a more comparative demonstration for the small energy test values.

Table 4 summarizes the best-fitted candidate for different speech features and frame lengths according to Figs. 1 and 2 evaluation results. According to Figs. 1 and 2 and Table 4, the following conclusions are conducted:

- The best-fitted candidate in the sense of the energy test for the T, RDFT, IDFT, and DCT features with frame length of 20, 30, and 100 ms is MLD, despite the often used assumption of multivariate Gaussian distribution in the speech enhancement algorithms [8–10], but consistent with the univariate Laplace distribution proposed by Martin [6] and Gazor et al. [14].
- The univariate Rayleigh distribution has been proposed for ADFT feature with a short frame length. Maybe as a consequence, it was expected that multivariate Rayleigh distribution (CRD) would be superior in modeling the multivariate distribution of ADFT; however, the energy test evaluation results proposed the CLD as the best-fitted candidate for frame lengths shorter than 500 ms.
- Regarding statistical modeling of ADFT features with short frame length, although CLD and CLID are both Laplace-based distribution, CLD was proposed as the best-fitted candidate. As the copula density function $c_X(.)$, which models inter-dimensional dependency, is non-unit for CLD and unit for CLID, the superiority of CLD over CLID shows how the modeling of inter-dimensional dependency contributes to the proper multivariate statistical modeling.
- Increasing frame length to 500 ms caused the best-fitted candidate corresponding to ADFT, RDFT, IDFT, and DCT features to be shifted from either CLD or MLD toward MGLD. This finding suggests that the Gaussian distribution contributed to the actual multivariate distribution of those domains when the frame length sufficiently increased, which is also supported by the central limit theorem. Similarly, varying the best-fitted distribution for LPC features from MGLD (with $p = 0.25$) to MGD verifies this contribution, too.
- The best-fitted candidate for the MFCC with different frame lengths is MGD, consistent with the assumption of multivariate Gaussian distribution used in most speech recognition algorithms [2, 3].

A: First best-candidate	B: Second best-candidate
C= Energy test value of A	D= Energy test value of B

Fig. 3 Values of cells in each block of Tables 5, 6, 7, 8, 9, 10, and 11

Table 5 Best-fitted multivariate distribution based on the energy test for ADFT coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.50$	MGLD, $p = 0.25$	CGevD	CGD
	0.05	0.06	0.03	0.04	0.02	0.02	0.00	0.01
Vowel	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.50$	CGevD	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.25$
	0.04	0.05	0.04	0.05	0.03	0.04	0.02	0.02
Nasal	MGLD, $p = 0.50$	CGevD	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGLD, $p = 0.50$
	0.04	0.04	0.02	0.02	0.03	0.03	0.00	0.00
Fricative	MGLD, $p = 0.75$	CGD	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$
	0.04	0.04	0.02	0.02	0.03	0.03	0.00	0.00
Stop	CGD	CGevD	CGD	CGevD	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.25$	MGLD, $p = 0.50$
	0.06	0.08	0.07	0.09	0.03	0.03	0.00	0.00

According to the conclusions, representative frames of speech signals containing T, RDFT, IDFT, or DCT features that are often used in statistical model-based speech enhancement algorithms [9, 10, 28] can be better statistically modeled by the MLD than MGD distribution. Furthermore, if the statistical-based algorithm exploits MFCC and ADFT, the energy test proposes MGD and ADFT, respectively.

Tables 5, 6, 7, 8, 9, 10, and 11 present the evaluation results of the energy test for the second experimental setup. In each table, there are 15 blocks surrounded by bold lines belonging to each phoneme class with a determined frame length. Each block in these tables contains four cells, as shown by Fig. 3. The A and B cells show the first and the second best-fitted candidates, respectively, and C and D cells indicate the energy test value corresponding to the first and the second best-fitted candidates, respectively.

According to Tables 5, 6, 7, 8, 9, 10, and 11, the following conclusions are conducted:

- The univariate Rayleigh distribution has been proposed for statistical univariate modeling of ADFT feature. Maybe as a consequence, it was expected that multivariate Rayleigh distribution (CRD) would be also superior in modeling multivariate distribution of ADFT; however, the evaluation results proposed MGLD, CGD, or CGevD as the best-fitted candidates.
- The best-fitted candidate in the sense of the energy test for all phoneme classes in T, RDFT, IDFT, and DCT features with different frame lengths was either MLD or MGLD (with $p \in \{0.25, 0.50, 0.75\}$). In particular for frame lengths of 20 and 30 ms, which are mostly exploited in speech processing, either MLD or MGLD with $p = 0.75$ dominated. As a

Table 6 Best-fitted multivariate distribution based on the energy test for LPC coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGD	MGLD, $p = 0.25$
	0.07	0.08	0.00	0.00	0.00	0.00	0.02	0.02
Vowel	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD
	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Nasal	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$
	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Fricative	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGD	CLD
	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03
Stop	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGLD, $p = 0.50$	MGLD, $p = 0.25$	MGD	MGD	MGLD, $p = 0.25$
	0.02	0.02	0.01	0.01	0.00	0.00	0.00	0.00

Table 7 Best-fitted multivariate distribution based on the energy test for time coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.25$	MGLD, $p = 0.50$
	0.01	0.02	0.01	0.02	0.00	0.01	0.00	0.00
Vowel	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MGLD, $p = 0.50$
	0.01	0.02	0.00	0.02	0.00	0.01	0.00	0.01
Nasal	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.50$	MGLD, $p = 0.50$	MLD	MGLD, $p = 0.25$
	0.00	0.01	0.01	0.02	0.00	0.00	0.00	0.01
Fricative	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.25$	MGLD, $p = 0.50$
	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Stop	MLD	CLD	MLD	CLD	MGLD, $p = 0.75$	MLD	MLD	MGLD, $p = 0.25$
	0.08	0.12	0.05	0.08	0.01	0.01	0.00	0.00

consequence, the Laplace distribution contributes more compared to the Gaussian distribution in the statistical multivariate modeling of T, RDFT, IDFT, and DCT features with short frame lengths.

- The best-fitted candidates for different phoneme classes with LPC feature was mostly MGD or MGLD with $p = 0.25$. As a consequence, the Gaussian distribution contributed more in the statistical multivariate modeling of LPC feature compared to the Laplace distribution.
- As the first or second best-fitted candidates for different process domains of a phoneme class with a fixed frame length mostly varied between MLD and MGLD (with $p \in \{0.25, 0.50, 0.75\}$), the statistical modeling of phonemes with a mixture of Gaussian and Laplace distributions is proposed.
- The best-fitted candidate for most phoneme classes with MFCC features extracted from frames of length less than 500 ms is MGD, consistent with the assumption of multivariate Gaussian distribution used in most speech recognition algorithms [2, 3].
- The only copula-based distribution proposed by the energy test evaluation results was CGD for statistical modeling of the stop phoneme class in ADFT domain with frame lengths of 20 and 30 ms, and CGevD for semivowel/glide with frame length of 500 ms.
- Based on the evaluation results, in the sense of the energy test, the copula-based distributions using IFM method were mostly overcome by conventional distributions in the second experimental setup. As only one of parameter estimation methods of copula-based distribution, IFM method, was taken into account in the experimental evaluation, and the IFM method ends up a sub-optimal solution for parameter estimation, it is difficult to have a generic conclusion on copula-based distribution's benefit in statistical modeling of speech frame.

Table 8 Best-fitted multivariate distribution based on the energy test for MFCC coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Vowel	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$
	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00
Nasal	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGLD, $p = 0.75$	MGLD, $p = 0.50$
	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01
Fricative	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	CLD	CLID
	0.00	0.01	0.00	0.00	0.00	0.00	0.03	0.03
Stop	MGD	MGLD, $p = 0.25$	MGD	MGLD, $p = 0.25$	MGLD, $p = 0.25$	MGD	CLD	MGD
	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01

Table 9 Best-fitted multivariate distribution based on the energy test for RDFT coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	CLD
	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00
Vowel	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MLD
	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Nasal	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	CGevD
	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01
Fricative	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.50$	MGD
	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Stop	MLD	CLID	MLD	CLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.25$	MLD
	0.06	0.10	0.04	0.08	0.01	0.01	0.00	0.00

One of future work perspective might therefore be to study the power of statistical modeling of copula-based distribution of speech frame using optimal parameter estimation methods.

– In some cases, where the energy test values of the first and second best-fitted candidates are almost the same, there is almost no superiority in the sense of energy test between the first or second best-fitted distributions, e.g., the case of fricative phoneme in time domain with different frame lengths in Table 7.

6 Conclusions

In this paper, the multivariate distribution of speech features in various domains, e.g., time, DFT, DCT, MFCC, and LPC, was studied and a framework was proposed for exploring the best-fitted distribution among different candidates. Ten plausible candidates including five conventional distributions, e.g., the multivariate Gaussian, multivariate Laplace, and the mixture of Gaussian–Laplace distributions (in three forms), and five copula-based distributions with marginal Laplace, independent marginal Laplace, Rayleigh, Gamma, and generalized extreme value (GEV) distributions were considered to explore the effect of feature type, phoneme class (for English language), and frame length on the distribution.

The evaluation results of the test energy showed that the multivariate Laplace distribution statistically better models time and DFT features of speech signals compared to the multivariate Gaussian distribution. For the amplitude of DFT features, the copula-based distribution with marginal Laplace distribution was proposed as the best-fitted candidate. For the MFCC features, the best-fitted candidate was MGD, consistent with the assumption of multivariate Gaussian distribution used in most speech recognition algorithms. For multivariate statistical modeling of different phoneme classes, the first or second best-fitted candidates for different domains (and also

Table 10 Best-fitted multivariate distribution based on the energy test for IDFT coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/glide	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	CGevD
	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00
Vowel	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	CLD
	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01
Nasal	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	CLD	MLD	CLD
	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01
Fricative	MGLD, $p = 0.75$	MLD	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.50$	CLID	MGLD, $p = 0.75$	CLD
	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00
Stop	MLD	CLD	MLD	CLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.50$	CLD
	0.08	0.12	0.04	0.08	0.01	0.01	0.00	0.01

Table 11 Best-fitted multivariate distribution based on the energy test for DCT coefficients of five phoneme classes in different frame lengths

Phoneme class	Frame length							
	20 ms		30 ms		100 ms		500 ms	
Semivowel/ glide	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD
	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
Vowel	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD
	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01
Nasal	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.50$	MGD	MGLD, $p = 0.50$
	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.02
Fricative	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.75$	MLD	MGLD, $p = 0.50$	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MGLD, $p = 0.50$
	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00
Stop	MLD	CLD	MLD	CLD	MLD	MGLD, $p = 0.75$	MGLD, $p = 0.75$	MGLD, $p = 0.50$
	0.05	0.10	0.04	0.08	0.01	0.01	0.00	0.00

for different frame sizes) mostly varied between MLD and MGLD (with $p \in \{0.25, 0.50, 0.75\}$), i.e., a mixture of Gaussian and Laplace distributions. The future work of this study can lead toward the development of statistical speech processing algorithms exploiting Laplace, mixture of Laplace and Gaussian, or copula-based multivariate distribution, depending on the feature type, phoneme class, and frame length.

Although the copula-based distribution was proposed as the best-fitted distribution for the modeling of amplitude of DFT, it is not the case for other features. It means that the copula-based approach requires more investigation in numbers of ways. First, the practical issues, e.g., the computational cost and the lack of sufficient amount of data for parameter estimation of some phoneme classes, e.g., stops, are needed to be considered. Second, as the IFM method used for parameter estimation of copula-based distribution ends up in a sub-optimal estimate, developing an optimal parameter estimation method for large vector dimensions is needed to have a fair evaluation of the copula-based distribution power in the statistical modeling of speech signals, e.g., compared to the optimal parameter estimation method used for MLD and MGD.

7 Appendix

The quantity ϕ, the energy, is defined as the difference between two pdfs $f_{X_0}(\boldsymbol{x})$ and $f_X(\boldsymbol{x})$ by

$$\begin{aligned}\phi &= \frac{1}{2}\int\int [f(x)-f_0(x)]\,[f(x')-f_0(x')]R(x,x')\,dx\,dx' \\ &= \frac{1}{2}\int\int [f(x)f(x') + f_0(x)f_0(x') - 2f(x)f_0(x')R(x,x')]\,dx\,dx'\end{aligned} \tag{28}$$

where the weight function R is a monotonically decreasing function of Euclidian distance and the integrals extend over the full variable space [44]. As the product of same distribution occurs in the first and second terms, it is not necessary to draw two different samples of the same pdf, and thereby, the first two terms can be neglected. The remaining third term has the form of expectation value of R and can be computed from the mean of all combinations $\boldsymbol{x}_{i=1}^{N} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^i, ..., \boldsymbol{x}^N\}$ following an unknown pdf $f(\boldsymbol{x})$ and simulated Monte–Carlo samples $\boldsymbol{q}_{j=1}^{N} = \{\boldsymbol{q}^1, ..., \boldsymbol{q}^j, ..., \boldsymbol{q}^M\}$ following $f_0(\boldsymbol{x})$, thus the energy statistic can be given by Eq. (29).

$$\begin{aligned}\phi_{NM} = &\frac{1}{N(N-1)}\sum_{t>i} R(|\boldsymbol{x}^i - \boldsymbol{x}^t|) \\ &+ \frac{1}{M(M-1)}\sum_{j>n} R(|q^n - q^j|) \\ &- \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} R(|\boldsymbol{x}^i - \boldsymbol{q}^j|)\end{aligned} \tag{29}$$

It is noted that since the evaluation of ϕ requires a summation over integrals, which is typically difficult, f_0 is preferred to be represented by a set of samples generated through a Monte–Carlo simulation.

8 Additional file

Additional file 1: List of speakers and sentences from TIMIT dataset used in the evaluations.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

In this paper, AA and HV developed the concept and designed experiments. Also, AA and HV prepared the first draft of the paper. AA implemented the algorithms and did the simulations. HS supervised the project and edited the manuscript. The final design of copula-based technique and also analysis of the related experiments were done by ZM. HV and AA prepared the revisions of the paper, and ZM edited the revisions. All authors discussed the final

results. HV is the corresponding author of the paper. All authors read and approved the final manuscript.

Author details
[1]Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany. [2]Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran. [3]Department of Computer Engineering, Sharif University of Technology, Tehran, Iran. [4]Department of Statistics, Shiraz University, Shiraz, Iran.

Received: 14 August 2014 Accepted: 29 November 2015
Published online: 30 December 2015

References

1. DY Zhao, Model based speech enhancement and coding, Ph.D., Royal Institute of Technology, KTH (2007)
2. L Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**, 257–286 (1989)
3. Huang, X, Acero, A, and Hon, H-W, Spoken language processing: a guide to theory, algorithm, and system development, Prentice Hall PTR, 2001
4. H Zen, K Tokuda, AW Black, Statistical parametric speech synthesis. Speech Comm. **51**, 1039–1064 (2009)
5. Z Xin, P Jancovic, L Ju, M Kokuer, Speech signal enhancement based on map algorithm in the ICA space. Signal Processing, IEEE Transactions on **56**, 1812–1820 (2008)
6. R Martin, Speech enhancement based on minimum mean-square error estimation and super-Gaussian priors. Speech and Audio Processing, IEEE Transactions on **13**, 845–856 (2005)
7. J-W Shin, J-H Chang, NS Kim, Statistical modeling of speech signals based on generalized Gamma distribution. Signal Processing Letters, IEEE **12**(3), 258–261 (2005)
8. A Aroudi, H Veisi, and H Sameti, Hidden Markov Model-based Speech Enhancement Using Multivariate Laplace and Gaussian Distributions, IET Signal Processing, **9**(2), 177–185, 2015, (doi:10.1049/iet-spr.2014.0032)
9. H Veisi, H Sameti, Speech enhancement using hidden Markov models in Mel-frequency domain. Speech Commun. **55**, 205–220 (2013)
10. Y Ephraim, A Bayesian estimation approach for speech enhancement using hidden Markov models. Trans. Sig. Proc. **40**, 725–735 (1992)
11. Y Ephraim, D Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. Acoustics, Speech and Signal Processing, IEEE Transactions on **32**, 1109–1121 (1984)
12. R McAulay, M Malpass, Speech enhancement using a soft-decision noise suppression filter. Acoustics, Speech and Signal Processing, IEEE Transactions on **28**, 137–145 (1980)
13. A Fazel, S Chakrabartty, An overview of statistical pattern recognition techniques for speaker verification. Circuits and Systems Magazine, IEEE **11**, 62–81 (2011)
14. S Gazor, Z Wei, Speech probability distribution. Signal Processing Letters, IEEE **10**, 204–207 (2003)
15. Allen, AO, Probability, statistics, and queueing theory with computer science applications, Academic Press, Inc., 1978
16. Papoulis, A, Probability, random variables and stochastic processes, McGraw-Hill Companies, 1991
17. B Chen, PC Loizou, A Laplacian-based MMSE estimator for speech enhancement. Speech Commun. **49**, 134–143 (2007)
18. JS Erkelens, J Jensen, R Heusdens, Speech enhancement based on rayleigh mixture modeling of speech spectral amplitude distributions, European Signal Proc. Conf. EUSIPCO, 2007
19. H Brehm, W Stammler, Description and generation of spherically invariant speech-model signals. Signal Process. **12**, 119–141 (1987)
20. JP LeBlanc, PL De Leon, Speech separation by kurtosis maximization, in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, 1998
21. J Jensen, I Batina, RC Hendriks, R Heusdens, A study of the distribution of time-domain speech samples and discrete fourier coefficients, Proc. IEEE First BENELUX/DSP Valley Signal Processing Symposium, 2005
22. C Schölzel, P Friederichs, Multivariate non-normally distributed random variables in climate research; introduction to the copula approach. Nonlin. Processes Geophys. **15**, 761–772 (2008)
23. T Ané, C Kharoubi, Dependence structure and risk measure. J. Bus. **76**, 411–438 (2003)
24. JC Rodriguez, Measuring financial contagion: a copula approach. Journal of Empirical Finance **14**, 401–423 (2007)
25. C Genest, M Gendron, M Bourdeau-Brien, The advent of copulas in finance. European Journal of Finance **15**, 609–618 (2009)
26. G Palombo, Multivariate goodness of fit procedures for unbinned data: an annotated bibliography, *arXiv*, 2011, 1102.2407v1.
27. AW Black, H Zen, K Tokuda, Statistical parametric speech synthesis, in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference*, 2007
28. S Srinivasan, J Samuelsson, WB Kleijn, Codebook-based Bayesian speech enhancement for nonstationary environments. Audio, Speech, and Language Processing, IEEE Transactions on **15**, 441–452 (2007)
29. A Sklar, Fonctions de Répartition à n Dimensions et Leurs Marges. Publications Inst. Statist. Univ. Paris **8**, 229–231 (1959)
30. RB Nelsen, *An introduction to copulas* (Springer, New York, 2006)
31. E Bouye, VD, A Nikeghbali, G Riboulet, and T Roncalli,, 'Copulas for finance: a reading guide and some applications', Working Paper. Groupe de Recherche Operationnelle, Credit Lyonnais, Available at http://ssrn.com/abstract=1032533, Nov. 2013.
32. Joe, H, Multivariate Models and Dependence Concepts, Chapman and Hall, 1997
33. W Hoeffding, Scale—invariant correlation theory, in *The Collected Works of Wassily Hoeffding*, ed. by NI Fisher, PK Sen (Springer, New York, 1994)
34. P Embrechts, F Lindskog, and A McNeil, 'Modelling Dependence with Copulas and Applications to Risk Management', *Handbook of Heavy Tailed Distributions in Finance, Rachev, S. (ed)*, Elsevier, 329–38 2001
35. X Chen, Y Fan, Estimation of copula-based semiparametric time series models, Vanderbilt University Department of Economics, 2004.
36. G Biau, M Wegkamp, A note on minimum distance estimation of copula densities. Statistics & Probability Letters **73**, 105–114 (2005)
37. BVM Mendes, EFL De Melo, RB Nelsen, Robust fits for copula models. Communications in Statistics: Simulation and Computation **36**, 997–1017 (2007)
38. Bowman, AW and Azzalini, A., Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations, Clarendon Press, 1997
39. J Yan, Enjoy the joy of copulas: with a package copula. J. Stat. Softw. **21**, 1–21 (2007)
40. Bellman, R.E, Adaptive control processes: a guided tour, Princeton University Press, 1961
41. PJ Clark, FC Evans, Generalization of a nearest neighbor measure of dispersion for use in K dimensions. Ecology **60**, 316–317 (1979)
42. KV Mardia, Measures of multivariate skewness and kurtosis with applications. Biometrika **57**, 519–530 (1970)
43. JH Friedman, LC Rafsky, Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. Ann. Stat. **7**, 697–717 (1979)
44. B Aslan, G Zech, Statistical energy as a tool for binning-free, multivariate goodness-of-fit tests, two-sample comparison and unfolding. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **537**, 626–636 (2005)
45. V Schmidt, E James, Gentle: random number generation and Monte Carlo methods. Metrika **64**, 251–252 (2006)
46. G Muraleedharan, C.G.S.a.C.L., Characteristic and Moment Generating Functions of Generalized Extreme Value Distribution, *Sea Level Rise, Coastal Engineering, Shorelines and Tides*, Nova Science Publishers, 269–276 2011.
47. Minka, TP, 'Estimating a Gamma distribution', http://research.microsoft.com/en-us/um/people/minka/papers/minka-gamma.pdf, Dec. 2012.
48. Paul Embrechts, CK, Thomas Mikosch, Modelling extremal events: for insurance and finance, Springer, 1997
49. JC Lagarias, JA Reeds, MH Wright, PE Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J. on Optimization **9**, 112–147 (1998)
50. K Fragiadakis, SG Meintanis, Goodness-of-fit tests for multivariate Laplace distributions. Math. Comput. Model. **53**, 769–779 (2011)
51. Garofolo, JS, Lamel, LF, Fisher, WM, Fiscus, JG, Pallett, DS, and Dahlgren, NL, Acoustic phonetic continuous speech corpus, NIST, 1993
52. Efron, B. and Tibshirani, R.J., An introduction to the Bootstrap, CRC press, 1994
53. B Aslan, and G Zech, *A New Class of Binning Free, Multivariate Goodness-of-Fit Tests: The Energy Tests*, arXiv preprint hep-ex/0203010, 2002