# Factor Analysis based Semantic Variability Compensation for Automatic Conversation Representation

*Mohamed Bouallegue*[†], *Mohamed Morchid*[†], *Richard Dufour*[†],
Driss Matrouf[†], *Georges Linarès*[†] *and Renato De Mori*[†‡]

[†]LIA, University of Avignon, France
[‡]McGill University, School of Computer Science, Montreal, Quebec, Canada

{firstname.lastname}@univ-avignon.fr, rdemori@cs.mcgill.ca

## Abstract

The main objective of this paper is to identify themes from dialogues of telephone conversations in a real-life customer care service. In this task, the word semantic variability contained in these conversations may impact the classification performance by retaining the noise in their vectorial representation. In this article, we propose an original method to compensate this semantic variability using the Factor Analysis (FA) paradigm, initially designed for speech processing tasks to compensate the acoustic variability, mainly in Speaker Verification (SV) and Automatic Speech Recognition (ASR). In our proposal, we used the FA paradigm to estimate the semantic variability as an additive component located in a subspace of low dimension (with respect to the super-vector space). This additive semantic variability is estimated in Factor Analysis model space. From this estimation, a specific vector transformation is obtained and is applied to vectors of dialogue representation. Experiments are reported using a corpus collected in the call center of the Paris Transportation Service. Results show the effectiveness of the proposed representation paradigm with a theme identification accuracy of 80.0%, showing a significant improvement with respect to previous results on the same corpus.

**Index Terms**: Human/Human conversation representation, Semantic variability, Factor analysis, Variability compensation, Automatic classification, Latent Dirichlet Allocation.

## 1. Introduction

Automatic Speech Recognition (ASR) systems globally achieve a sufficient level of performance to be used in various tasks, such as text analysis, automatic classification or information extraction. Nonetheless, particular speech conditions (conversational speech, noisy environments...) may drastically drop the transcription accuracy, which directly affects the applications based on the ASR outputs. Two ways are typically followed to deal with these speech recognition errors. The first one consists in adapting ASR systems to a targeted domain and specific conditions. As a result, the transcription accuracy should increase, but such an approach usually requires task-specific speech materials and costly manual annotations. The second way is not to seek to correct these errors, but to propose additional solutions to compensate them. This last solution is the one followed in this article.

This work was funded by the SUMACC and ContNomina projects supported by the French National Research Agency (ANR) under contracts ANR-10-CORD-007 and ANR-12-BS02-0009.

This solution estimates a topic space, with, for example, a Latent Dirichlet Allocation (LDA) approach [1], in which each document may be viewed as a mixture of latent topics. Nonetheless, this projection of noisy documents into a clean topic space generates a variability (called *semantic variability*) in the dialogue vectorial representation. This variability is mainly due to the few number of words contained in each dialogue. Although the effect of this semantic variability is limited compared to the usually noted gain provided by this abstracted content representation, the semantic variability may degrade the dialogue representation.

In this paper, we will call *semantic variability* all kinds of variability affecting the vectorial representation of given documents. We propose a new method to compensate the semantic variability in order to obtain a better dialogue representation. Our proposal is based on the Factor Analysis (FA) paradigm presented in [2]. This approach was applied in the speaker recognition domain to model the session variability[1] as an additive component. The basic idea behind this approach consists in decomposing the speaker model into three different components: a speaker-session independent component, a speaker dependent component, and a session dependent component. This decomposition allows to easily remove the session dependent component which degrades the performance of speaker verification systems.

We will apply the same paradigm (FA) to the theme identification problem. The dialogues being considered as the speakers in the FA, and the semantic variability replacing the session variability. The FA approach is used to estimate the component of this variability in order to compensate it in the dialogue representation. The neutralization of this variability should allow to obtain a more robust dialogue vectorial representation.

This paper is organized as follows. The dialogue representation is described in Section 2. In Section 3, the proposed approach to model the theme of the dialogue by using the SGMM is presented. Sections 4 and 5 report experimental results, while Section 6 concludes this work.

## 2. Dialogue vectorial representation

The considered application is the identification of the major theme of a human/human telephone conversation in the customer care service of the RATP Paris transportation system. The approach considered in this paper focuses on modeling the vari-

[1]This term encompasses a number of phenomena including transmission channel effects, transducer characteristics, environment noise, and variability introduced by the speaker.

ability between different dialogues expressing the same theme $C$. For this purpose, it is important to select features that represent semantic contents relevant for the theme of a dialogue. An attractive set of features for capturing possible semantically relevant word dependencies is obtained with Latent Dirichlet Allocation (LDA) [1], a generative probabilistic model.

A dialogue is then represented as a finite mixture over an underlying set of topics. Given a train set of conversations, a hidden topic space is derived and a conversation $d$ is represented by its probability in each topic of the hidden space. Estimation of these probabilities is affected by a variability inherent to the estimation of the model parameters. If many hidden spaces are considered and features are computed for each hidden space, it is possible to model the estimation variability together with the variability of the linguistic expression of a theme by different speakers in different real-life situations.

This multiple representation of a dialogue, even if the purpose of the application is theme identification and a train corpus annotated with themes is available, supervised LDA [3] is not suitable for the proposed approach since LDA is used only for producing different feature sets used for computing statistical variability models.

To estimate the parameters of different hidden spaces, a vocabulary $V$ of discriminative words is constructed as described in [4, 5, 6]. For each theme $\{C_i\}_{i=1}^{8}$, a set of 50 theme specific words is identified. The same word may appear in more than one theme vocabulary selection. All the selected words are then merged without repetition to form $V$ made of 166 words.

Several techniques have been proposed to estimate the LDA parameters, such as Variational Methods [1], Expectation Propagation [7], or Gibbs Sampling [3, 8]. Gibbs Sampling is a special case of Markov-chain Monte Carlo (MCMC) [9] and gives a simple algorithm for approximate inference in high-dimensional models such as LDA [8]. This overcomes the difficulty to directly and exactly estimate parameters that maximize the likelihood of the whole data collection defined as: $P(W|\overrightarrow{\alpha}, \overrightarrow{\beta}) = \prod_{\overrightarrow{w} \in W} P(\overrightarrow{w}|\overrightarrow{\alpha}, \overrightarrow{\beta})$ for the whole data collection $W$ knowing the Dirichlet parameters $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$.

The Gibbs Sampling allows us both to estimate the LDA parameters, to represent a new dialogue $d$ with the $n^{\text{th}}$ topic space $\Gamma_n^q$ of size $q$, and to obtain a feature vector $V_d^{z^n}$ of the topic representation of $d$. The $k^{th}$ feature $V_d^{z_k^n} = P(z_k^n|d)$ (where $1 \leq k \leq q$) is the probability of topic $z_k^n$ is generated by the unseen dialogue $d$ in the $n^{\text{th}}$ topic space of size $q$ and $V_{z_k^n}^{w_i} = P(w_i|z_k^n)$ is the vector representation of a word $w_i$ into $\Gamma_n^q$.

In the LDA technique, the topic $z$ is drawn from a multinomial over $\theta$ which is drawn from a Dirichlet distribution over $\overrightarrow{\alpha}$. Thus, a set of $p$ topic spaces $\{\Gamma_n^q\}_{n=1}^{p}$ of size $q$ are learned using LDA by varying the topic distribution parameter $\overrightarrow{\alpha} = [\alpha_1, \ldots, \alpha_q]^t$ to obtain $p$ topic spaces of size $q$. The standard heuristic is $\alpha_i = \frac{50}{q}$ [3], which for the setup of the $n^{\text{th}}$ topic space ($1 \leq n \leq p$) would be $\overrightarrow{\alpha_n}[\underbrace{\alpha_n, \ldots, \alpha_n}_{\text{q times}}]^t$ with $\alpha_n = \frac{n}{p} \times \frac{50}{q}$.

The larger $\alpha_n$ ($\alpha_n \geq 1$) is, the more uniform $P(z|d)$ will be. Nonetheless, this is not what we want: different dialogues have to be associated with different topic distributions. In the meantime, the higher the $\alpha$ is, the more the draws from the Dirichlet will be concentrated around the mean, which, for a symmetric alpha vector, will be the uniform distribution over $q$. The number of topics $q$ is fixed to 50, and 500 topic spaces are built ($p = 500$) in our experiments. Thus, $\alpha_n$ varies between a low value (sparse topic distribution $\alpha_1 = 0.002$) to 1 (uniform Dirichlet $\alpha_q = 1$).

The next process allows to obtain a homogeneous representation of the dialogue $d$ for the $n^{\text{th}}$ topic space $\Gamma_n^q$. The feature vector $V_d^{z^n}$ of the dialogue $d$ is mapped into the common vocabulary space $V$ composed with a set of discriminative words [4, 5, 6] to obtain a new feature vector [10] $V_{d,n}^{w} = [P(w|d)_{\Gamma_n^q}]_{w \in V}$ of size 166 for the $n^{\text{th}}$ topic space $\Gamma_n^q$ of size $q$ where the $i^{\text{th}}$ ($0 \leq i \leq 166$) feature is:

$$V_{d,n}^{w_i} = \sum_{k=1}^{q} P(w_i|z_k^n) P(z_k^n|d) = \sum_{k=1}^{q} V_{z_k^n}^{w_i} \times V_d^{z_k^n}$$

# 3. Factor analysis for semantic variability compensation

In this section, we present a new method of semantic variability compensation of automatically transcribed dialogues. Our idea is inspired from the acoustic variability compensation successfully applied for speech processing (speech recognition [11], speaker verification [12], and language identification [13]) based on the Factor Analysis (FA) paradigm. The basic idea of this approach is to project a vector representing a noisy speech in a subspace assumed to only contain the noise part. Hence, this projected component can be subtracted from the noisy vector to obtain a clean speech vector.

## 3.1. Compensation of nuisance variability for speech processing

In the context of speech processing, the FA process is performed in the cepstral domain, assuming that the whole cepstral space is modeled by a Gaussian Mixture Model, called *Universal Background Model* (UBM). The useful information $i$ (*speaker* in the speaker verification context [12], *phoneme* in the ASR case [11]) can be modeled by a GMM derived from the UBM using a MAP adaptation of the UBM vector means. The concatenation of the GMM means allows to obtain a very high dimensional vector, called a *super-vector*.

The FA paradigm gives the possibility to model the useless information $h$ (environment variability, background noise, speaker-variability, channel-variability...) in a subspace of low dimension $R$, in order to remove it from the noisy super-vector. Let $M$ be the number of Gaussians in the GMM-UBM. The GMM-UBM is trained on a large amount of data. Let $\mathbf{m}$ be the super-vector (of dimension $MD$ where $\mathbf{D}$ is the dimension of the acoustic space) obtained by the concatenation of all means in GMM-UBM. By using the FA paradigm, the super-vector $\mathbf{m}_{i,h}$ (random variable) can be decomposed into three different components:

$$\mathbf{m}_{h,i} = \mathbf{m} + \mathbf{D}\mathbf{y}_i + \mathbf{U}\mathbf{x}_{h,i} \quad (1)$$

In details, $\mathbf{y}_i$ models the useful information, which is a vector of dimension $MD$, and $\mathbf{Ux}$ is the nuisance variability component. $\mathbf{x}_h$ are the nuisance variability factors (vector of dimension $R$). Both $\mathbf{y}_i$ and $\mathbf{x}_h$ are assumed to be normally distributed among $N(0, I)$. $\mathbf{D}$ is a diagonal matrix ($MD \times MD$) so that $\mathbf{DD}^t$ is the *a priori* covariance matrix of the useful component. $\mathbf{D}$ satisfies the equation $I = \tau D^t \Sigma^{-1} D$, where $\tau$ is the relevance factor required in the standard MAP adaptation. $\mathbf{U}$ is a rectangular matrix ($MD \times R$) so that $\mathbf{UU}^t$ is the *a priori* covariance matrix of the nuisance variability component random

vector. The algorithm that presents the adopted strategy to estimate different components of the equation 1 is detailed in [14].

As shown in equation 1, the success of the Factor Analysis modeling mainly depends on the assumptions that the nuisance variability is located in a subspace of low dimension (dimension $R$) and that the nuisance variability is additive ($\mathbf{U}\mathbf{x}_h$). Once the different component of the equation 1 is finite. The nuisance variability component **Ux** is used in subtracting of this variability from the vector dialogue representation from trains and test data.

### 3.2. Semantic variability compensation for automatic theme identification task

In this work, we apply the Factor Analysis paradigm in the context of the automatic theme identification. In this task, ASR transcriptions are used. Since the transcription quality may be poor, we propose to represent the dialogue in a more abstract representation using a topic space. This abstract representation of the dialogue spoken content allows to limit the negative impact of ASR errors. Nonetheless, the projection of the dialogues in a topic space generates a variability due to the estimation of the LDA parameters. Their estimation with the Gibbs Sampling [15] takes into account all words contained in the vocabulary. Indeed, for an unseen dialogue $d$, the estimation of the probability that a topic $z$ was generated by $d$ adds a residual semantic variability due to the fact that $p(z|d)$ is estimated for all words in the vocabulary, and not only for the words contained in $d$. This variability degrades the quality of the dialogue representation which could affect the theme identification performance. For this reason we propose to use the Factor Analysis approach to estimate and remove this variability in order to improve the quality of the dialogue representation.

#### *3.2.1. Semantic variability estimation*

At this stage, the dialogues are represented by vectors. A GMM-UBM is firstly estimated using a train corpus (a set of dialogue vectors). This model is defined as follows:

$$UBM = (\alpha_g, m_g, \Sigma_g) \tag{2}$$

where $\alpha_g$, $m_g$ and $\Sigma_g$ are respectively the weight, the mean and the covariance matrix of the $g^{th}$ Gaussian. The GMM-UBM models the set of training dialogues with the semantic variability contained in these dialogues.

Let $N$ be the dimension of the dialogue representation (vectorial representation), $M$ is the number of Gaussians in the GMM-UBM, and $R$ is the chosen dimension of the semantic variability subspace. Let $v$ be the semantic variability in the dialogue representation $d$. The super-vector $m_{v,d}$ of this dialogue in the presence of variability $v$ can be obtained using the following equation:

$$\mathbf{m}_{v,d} = \mathbf{m} + \mathbf{D}\mathbf{y}_d + \mathbf{U}\mathbf{x}_{v,d} \tag{3}$$

where **m** is the super-vector of the GMM-UBM. **Ux** is the semantic variability component. The columns of the **U** (a $MD \times R$ matrix) are the generative vectors of the semantic variability. $\mathbf{x}_{(v,d)}$ is the semantic variability vector in the subspace generated by the columns of the matrix **U**. The vector $\mathbf{y}_{ph}$ models the dialogue $d$ and **D** is a $MD \times M$ D diagonal matrix. The algorithm to estimate the different components is detailed in [14].

The variability component **Ux** have to be compensated in the vectorial dialogues representation. In next section, we present the way that this component is subtracted from the differants vectors dialogue representation.

#### *3.2.2. Semantic variability compensation*

In this step, we use the **U** matrix, estimated in the last section, to compensate the semantic variability $v$ from the dialogue representation. The same matrix **U** was used for all dialogue representations contained in the train and test corpus. The clean dialogue representation $\hat{R_d}$ is obtained by using the following equation:

$$\hat{R_d} = R_d - \sum_{g=1}^{M} \gamma_g(t) \cdot \{\mathbf{U} \cdot \mathbf{x}_{v,d}\}_{[g]} \tag{4}$$

where $M$ is the number of Gaussians in the GMM-UBM, and $\gamma_g(t)$ is the *a posteriori* probability of Gaussian $g$ given by the dialogue representation $R_d$. These probabilities are estimated by using the GMM-UBM model. $\mathbf{U}\mathbf{x}_{v,d}$ is the additive semantic variability component estimated on the original dialogue representations. The clean dialogue representation obtained $\hat{R_d}$ will be evaluated in the context of theme identification.

## 4. Experimental protocol in the theme classification task

In the previous section, we proposed a method to compensate the semantic variability contained in the dialogue vector representation. We present in this section the corpus used in our experiments, as well as the classical Mahalanobis distance and EFR standardization approach that will be used in the classification of the topic-based dialogue representation.

### 4.1. The DECODA project

The experiments on theme identification are performed using the DECODA project corpus [16]. This corpus is composed of 1,067 telephone conversations split into a train set (740 dialogues) and a test set (327 dialogues), and manually annotated with 8 conversation themes: *problems of itinerary*, *lost and found*, *time schedules*, *transportation cards*, *state of the traffic*, *fares*, *infractions* and *special offers*. The set of 500 topic spaces needed for these experiments (see Section 2), is built with the use of Mallet Java implementation of LDA[2].

The Automatic Speech Recognition (ASR) system used for the experiments is LIA-Speeral [17] with 230,000 Gaussians in the triphone acoustic models. Model parameters were estimated with maximum *a posteriori* probability (MAP) adaptation from 150 hours of speech in telephone condition. The vocabulary contains 5,782 words. A 3-gram language model (LM) was obtained by adapting with the transcriptions of the train set a basic LM. An initial set of experiments was performed with this system resulting in an overall Word Error Rate (WER) on the train set of 45.8% and on the test set of 58.0%. These high WER are mainly due to speech disfluencies and to adverse acoustic environments for some dialogues when, for example, users are calling from train stations or noisy streets with mobile phones. Furthermore, the signal of some sentences is saturated or of low intensity. A "stop list" of 126 words[3] was used to remove unnecessary words which results in a WER of 33.8% on the train set and of 49.5% on the test set.

[2] http://mallet.cs.umass.edu/

[3] http://code.google.com/p/stop-words/

### 4.2. Mahalanobis distance

Given a new observation $x$, the goal of the task is to identify the theme belonging to $x$. The probabilistic approaches ignore the process by which vectors were extracted, and they pretend instead they were generated by a prescribed generative model. The representation mechanism of the dialogue is ignored and is regarded as an observation from a probabilistic generative model. The two most simple assumptions are those of the homoscedastic Gaussian Bayesian classifier [18]: (i) the Gaussianity of the theme classes and (ii) the equality of the class covariances.

The Gaussian classifier is based on the Bayes decision rule and is combined with a scoring metric to assign a dialogue $d$ with the most likely theme $t$. Given a training dataset of dialogues $D$, let $\mathbf{W}$ denote the within dialogue covariance matrix defined by:

$$\mathbf{W} = \sum_{k=1}^{K} \frac{n_t}{n} \mathbf{W_k} = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=0}^{n_t} \left(x_i^k - \overline{x_k}\right) \left(x_i^k - \overline{x_k}\right)^t \quad (5)$$

where $\mathbf{W_k}$ is the covariance matrix of the $k^{\text{th}}$ theme $C_k$, $n_t$ is the number of utterances for the theme $k$, $n$ is the total number of dialogues in the training dataset, and $\overline{x_k}$ is the mean of all dialogues $x_i^k$ of the $k^{\text{th}}$ theme.

Each dialogue does not contribute to the covariance in an equivalent way. For this reason, the term $\frac{n_t}{n}$ is introduced in equation 5.

If homoscedasticity (equality of the class covariances) and Gaussian conditional density models are assumed, a new observation $x$ from the test data can be assigned to the most likely theme $k_{\text{Bayes}}$ using the Gaussian classifier based on the Bayes decision rule:

$$\begin{aligned} k_{\text{Bayes}} &= \arg\max_k \ \mathcal{N}\left(x \mid \overline{x_k}, \mathbf{W}\right) \\ &= \arg\max_k \left\{ -\frac{1}{2} \left(x - \overline{x_k}\right)^t \mathbf{W}^{-1} \left(x - \overline{x_k}\right) + a_k \right\} \end{aligned} \quad (6)$$

where $\overline{x_k}$ is the centroid (mean) of theme $k$, $\mathbf{W}$ is the within theme covariance matrix defined in equation 5, $\mathcal{N}$ denotes the normal distribution and $a_k$ is the log prior probability of the theme membership defined as $a_k = \log\left(P(C_k)\right)$. It is worth noting that, with these assumptions, the Bayesian approach is similar to the Fisher's geometric approach: $x$ is assigned to the nearest centroid's class, according to the Mahalanobis [19] metric of $\mathbf{W}^{-1}$:

$$k_{\text{Bayes}} = \arg\max_k \left\{ -\frac{1}{2} ||x - \overline{x_k}||^2_{\mathbf{W}^{-1}} + a_k \right\} \quad (7)$$

### 4.3. Vector standardization

Random variability have to be theoretically normally distributed among $\mathcal{N}(0, I)$, the vectors are standardized. Two methods showed improvements for speaker verification: Within Class Covariance Normalization (WCCN) [12] and Eigen Factor Radial (EFR) [20]. This last method includes length normalization [21]. Both of these methods dilate the total variability space as the mean to reduce the within-class variability. In this paper, the EFR technique is chosen to standardize the dialogue representation. The standardization operation is as follow:

$$\frac{x}{x' \Sigma x} \quad (8)$$

Where $\Sigma$ is the covariance matrix of the random variable $x$.

## 5. Results

Experiments are conducted using the multiple topic spaces estimated with a LDA approach (various number of topic spaces). From these multiple topic space configurations, the classical approach is to find the one that allows to reach the best theme classification performance. The proposed variability compensation method is applied in each topic-based representation of a dialogue and is compared with dialogue representation without variability compensation. The compensation is applied with different ranks $R$ of the matrix $U$ and different numbers of Gaussian $M$ in the GMM-UBM. In our experiments, the mean Mahalanobis score is computed as well as the EFR standardization in each case (filtered and non-filtered).

First experiments were made using the non-filtered vector representation of dialogues. This standard representation allowed to obtain an average accuracy[4] of $76.9\%$, which constitutes our baseline [4]. We then computed the average score obtained by the compensated vector representation, using different configurations (see Table 1). All of the average scores reported in Table 1 outperform our baseline result. We can point out that the best average score is obtained using 16 Gaussians in the GMM-UBM model and a $rank(U) = 60$ configuration, with a gain of 3.3 points in comparison to the basic vector representation. Finally, we can conclude that Factor Analysis, applied in the context of text classification, can significantly reduce the vector size while improving the classification results.

| | **Number of Gaussians in GMM-UBM** | | |
|---|---|---|---|
| Rank of **U** | *16* | *32* | *64* |
| *60* | **80.0** | 77.2 | 78.2 |
| *80* | 79.7 | 78.7 | 78.2 |
| *100* | 79.2 | 78.2 | 78.0 |

Table 1: Theme classification average accuracy (%) with compensated variability vectors.

## 6. Conclusions

In this paper, we proposed to apply a variability compensation approach, originally designed for speaker recognition problems, to improve the vectorial conversation representation. Indeed, the Factor Analysis space representation has never been applied to textual content, such as words or topic-based representation features. This compensated representation could then be applied to any task that uses textual vectorial representation.

We applied this approach to a theme identification classification task, and showed that this method allows to remove the low residual variability (semantic variability). As a consequence, results have been significantly improved in terms of average accuracy among imperfect transcriptions. A final gain of 3.1 points has then been noticed. These encouraging results give some good reasons to continue the adaptation of speech processing techniques into the natural language processing field such as document categorization, keywords extraction.... In a future work, we will use these vector representations to build more robust textual document representations.

[4] Average score obtained using the different topic space configurations (see section 4)

## 7. References


[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[2] Patrick Kenny, Najim Dehak, Vishwa Gupta, and P Dumouchel, "A new training regimen for factor analysis of speaker variability," *ICASSP*, 2008.

[3] Thomas L Griffiths and Mark Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.

[4] Mohamed Morchid, Georges Linarès, Marc El-Beze, and Renato De Mori, "Theme identification in telephone service conversations using quaternions of speech features," in *INTERSPEECH*, 2013.

[5] Mohamed Morchid, Richard Dufour, Pierre-Michel Bousquet, Mohamed Bouallegue, Georges Linarès, and Renato De Mori, "Improving dialogue classification using a topic space representation and a gaussian classifier based on the decision rule," in *ICASSP*, 2014.

[6] Mohamed Morchid, Richard Dufour, and Georges Linarès, "A LDA-based topic classification approach from highly imperfect automatic transcriptions," in *LREC'14*, 2014.

[7] Thomas Minka and John Lafferty, "Expectation-propagation for the generative aspect model," in *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 352–359.

[8] Gregor Heinrich, "Parameter estimation for text analysis," *Web: http://www. arbylon. net/publications/text-est. pdf*, 2005.

[9] Stuart Geman and Donald Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 6, pp. 721–741, 1984.

[10] M. Morchid, R. Dufour, and G. Linarès, "Thematic representation of short text messages with latent topics: Application in the twitter context," in *PACLING*, 2013.

[11] Mohamed Bouallegue, Mickael Rouvier, Driss Matrouf, and Georges Linares, "Noise compensation for speech recognition using subspace gaussian mixture models," *INTERSPEECH*, 2012.

[12] Najim Dehak, Patrick J Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[13] Florian Verdet, Driss Matrouf, Jean-Fran**c**cois Bonastre, and Jean Hennebert, "Factor analysis and svm for language recognition," in *INTERSPEECH*, 2009, pp. 164–167.

[14] Driss Matrouf, Nicolas Scheffer, Benoit G. B. Fauve, and Jean-Fran**c**cois Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification," in *INTERSPEECH*, 2007, pp. 1242–1245.

[15] James G Scott and Jason Baldridge, "A recursive estimate for the predictive likelihood in a topic model," .

[16] Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, and Eric Arbillot, "Decoda: a call-centre human-human spoken conversation corpus," LREC'12, 2012.

[17] Georges Linarès, Pascal Nocéra, Dominique Massonie, and Driss Matrouf, "The lia speech recognition system: from 10xrt to 1xrt," in *Text, Speech and Dialogue*. Springer, 2007, pp. 302–308.

[18] Sergios Petridis and Stavros J Perantonis, "On the relation between discriminant analysis and mutual information for supervised linear feature extraction," *Pattern Recognition*, vol. 37, no. 5, pp. 857–874, 2004.

[19] Eric P Xing, Michael I Jordan, Stuart Russell, and Andrew Ng, "Distance metric learning with application to clustering with side-information," in *Advances in neural information processing systems*, 2002, pp. 505–512.

[20] Pierre-Michel Bousquet, Driss Matrouf, and Jean-Fran**c**cois Bonastre, "Intersession compensation and scoring methods in the i-vectors space for speaker recognition.," in *INTERSPEECH*, 2011, pp. 485–488.

[21] Daniel Garcia-Romero and Carol Y Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems.," in *INTERSPEECH*, 2011, pp. 249–252.