A Comparison of Classification Paradigms for Speaker Likeability Determination

Nicholas Cummins[1], *Julien Epps*[1], *Jia Min Karen Kua*[1]

[1] School of Electrical Engineering and Telecommunications
The University of New South Wales, Sydney, NSW 2052, Australia

n.p.cummins@unsw.edu.au, j.epps@unsw.edu.au

Abstract

In this paper we investigate the performance of different classification paradigms, testing each with a range of acoustic features, to find a system that is well suited to speaker likeability classification. We introduce a Sparse Representation Classifier for paralinguistic classification and explore the role of training data selection for a GMM classifier. Results demonstrate that (1) Single dimensional features of pitch direction, shimmer and spectral roll-off were the most suitable features found when testing on the development set but we were unable to reproduce their performance in the final classification task, (2) Using UBM training data selection increased accuracy of MFCC's and (3) Sparse Representation showed promise as a paralinguistic classifier with results comparable to that of SVM.

Index Terms: Likeability, single and multi-dimensional feature selection, Sparse Representation Classifier, UBM data selection.

1. Introduction

The paralinguistic component of speech communication contains information that can be used to infer a wide range of information about the speaker. Previous Interspeech challenges have targeted traditional paralinguistic fields of emotions, age, gender or speaker states such as sleepiness, intoxication. The 2012 challenge focuses on more abstract speaker traits including personality, likeability and pathology. Herein, we describe a system and related experiments submitted to the likeability sub-challenge.

The concept of likeability is difficult to define; popular definitions often include adjectives such as ‘agreeable’, ‘pleasant’ and ‘easy to like’. For this paper the definition of likeable is set by the constraints of the database and its annotation. The database is a subset of the Agender database, used in the 2010 paralinguistic challenge. The likeability ratings were established by listeners ranking utterances on a seven point Likert scale, the results of which were discretised into ‘likeable’ and ‘not likeable’ classes - for further information see [1].

There has been some recent work examining the effect of social behaviours such as friendliness, flirtiness and personality types such as extroversion and pleasantness on speech. These traits have analogies with likeability. Mairesse et al. link extroversion to positive affect and show that prosodic features including high speech rate, greater vocal intensity and high pitch variability can be used to describe extroverted speech [2]. Jurafsky, Ranganath and McFarland analysed recordings taken during speed dating and found that men identified as friendly voices that had lower intensity and a lower and less varied pitch. Women who were identified as friendly had voices with a higher pitch but lower and less varied intensity [3]. Recently Pinto-Coelho et al. used a wide range of prosodic and voice quality features to identify pleasantness [4].

In the last two years there have been a small number of papers trying to find correlates between acoustic parameters and likeability. Weiss and Burkhardt showed that both prosodic and spectral features can be used to assess speaker likeability [5]. They showed that stereotypically likeable males have lower, deeper voices, which can be reflected in pitch patterns. Likeable women have bright, youthful sounding voices with energy spread over the spectrum. Using a binary decision tree classifier, an accuracy of 62.9% in likeability classification was achieved [5], albeit on a smaller database (10 speakers). Gravano et al. found that likeable speech from both genders had higher intensity, lower shimmer and lower pitch with a more reduced range [6]. They also showed that males, when trying to sound likeable, will lower their vocal intensity and expand pitch range. On the other hand, women will raise their pitch when trying to be liked by a male and lower pitch and increase intensity when speaking to a female.

A benchmark has been set by the challenge organisers of an unweighted binary classification accuracy of 58.5% using a support vector classifier [7]. A higher classifier accuracy was obtained by an initial investigation [1], which was able to achieve an unweighted classification accuracy of 67.6% using a binary tree classifier. Their work also shows that spectral, prosodic and voice quality features are suited to the classification task, whilst cepstral features, used on their own, give classification near chance level.

The 2012 likeability challenge is to the classify a speaker into either a likeable class or a non-likeable class through the use of acoustic features and a learning algorithm [7]. This involves finding acoustic patterns that correlate well with the subjective listener rankings given by the challenge organizers. Our approach to this task is to trial five different gender dependent feature/classifier systems, applying each to both baseline openSMILE features and selected other spectral features, with a view to outperforming the Challenge benchmark set and the classifier accuracy measured in [1].

2. Feature Extraction

2.1 Voice Activity Detection

To isolate the voiced sections of the likeability database we used openSMILE’s voiced probability measure, which employs the autocorrelation function and cepstrum field to calculate the probability of a frame representing voiced speech. Unless specifically stated, the voicing cut-off we used was the default setting of 55%.

2.2 Single-dimensional Feature Extraction

The Single dimensional features we used in our testing were extracted using openSMILE[1], and included prosodic measures such as frequency (F0), intensity and pitch direction score (F0ds) which represents whether the pitch of a pseudo syllable is falling, flat, or rising. Voice Quality measures such as jitter and shimmer were also extracted and a wide range of broad spectral measures such as zero crossing rate (ZCR), spectral bands and roll-off coefficients. All pitch based features were extracted using a 40ms frame, extracted every 10ms, whilst all other features used a 20ms frame extracted every 10ms. Work reported in [1, 5, 6] shows the potential of prosodic, voice quality and spectral measures features in likeability classification.

1. *http://opensmile.sourceforge.net/*

2.3 Multi-dimensional Feature Extraction

A range of detailed cepstrum and spectral features were tested for their suitability to the classification task. These include the mel frequency cepstral coefficients (MFCC), perceptual linear predictive cepstral coefficients (PLP) and line spectral pairs (LSP), all extracted using openSMILE. We also tested linear prediction cepstral coefficients (LPCC) and the spectral centroid frequencies and amplitudes (SCF / SCA). The LPCC were extracted using VoiceBox[2], and the reader is referred to [8] for the extraction method of the spectral centroid measures. All features were extracted using a 20ms frame with 10ms overlap. Work reported in [1] shows that detailed spectral features are potentially suitable for likeability classification whilst [5] shows the potential of the cepstral features.

3. Classification Approach and Experimental Setup

3.1 Classification Systems

Three different classifiers were trialed during our initial testing; Gaussian Mixture Models (GMM) and Support Vector Machines (SVM), together with a newer system, Sparse Representation Classification (SRC). The aim of this testing is to find the most suitable learning algorithm to use in our final system configuration.

Both GMM's and SVM's are popular classification techniques and have been shown to work in many different paralinguistic classification tasks so should be suitable for likeability classification. SRC is a newer classification approach in speaker recognition and can be thought of as somewhat complementary to an SVM classifier [9]. We were motivated to use the challenge as an opportunity to compare the performance of SRC as a paralinguistic classification method against a range of different learning systems. Section 3.2 gives a brief outline of the sparse classification process.

Given the results in [5, 6] we hypothesize that males and females encode likeability differently, so we train and test with separate models for both genders throughout.

3.2 Sparse Representation Classification

In recent years, sparse representation based classifiers have begun to emerge for various applications, and experimental results indicate that they can achieve comparable or better performance than other classifiers [9, 10]. Motivated by [9], herein we propose the use of vector–based (a vector is composed of either utterance statistics or supervectors) SRC for paralinguistic classification, which is formulated as follows. First, a matrix $\mathbf{A} \in \mathbb{R}^{K\text{x}N}$ is defined for the entire training set as the concatenation of n K-dimensional likeable (L) training samples and m K-dimensional not-likeable (NL) training samples:

$$\mathbf{A} = [\mathbf{A}_L, \mathbf{A}_{NL}] = \left[v_{L,1}, v_{L,2}, \dots, v_{L,n}, v_{NL,1}, \dots, v_{NL,m}\right] \quad (1)$$

where $N = n + m$. Then, the test sample $\mathbf{S} \in \mathbb{R}^{K}$ can be represented as a linear combination of all training samples as

$$\mathbf{s} = \mathbf{Ay} \quad (2)$$

Although $N > K$, which corresponds to an underdetermined system, its unique sparsest solution, $\mathbf{y}$, can be solved by ℓ_1-norm minimization [11]. Finally for classification, a characteristic function, $\delta_i\colon \mathbb{R}^N \longrightarrow \mathbb{R}^N$, that selects the coefficients associated with the ith class as shown in (3) for each class i (L or NL) is defined [10].

$$\delta_i(\mathbf{y}) = \left[\sigma_{L,1}\sigma_{L,2} \dots \sigma_{L,n} \dots \sigma_{NL,1}\sigma_{NL,2} \dots \sigma_{NL,m}\right]^T$$
$$where\ \sigma_{j,k} = \begin{cases} 0\ , j \notin class\ i\ \forall\ k \\ \alpha_{i,k}\ , j \in class\ i\ \forall\ k \end{cases} \quad (3)$$

Using only the coefficients associated with the ith class, the given test vector $\mathbf{S}$ can be approximated as $\widehat{\mathbf{s}_i} = \mathbf{A}\delta_i(\mathbf{y})$, $\mathbf{s}$ was then assigned to the class $\mathbb{C}_\mathbf{S}$ that gave the smallest residual between $\mathbf{s}$ and $\widehat{\mathbf{s}_i}$:

$$\mathbb{C}_\mathbf{S} = \arg\min_i r_i(\mathbf{s}) \quad where \quad r_i(\mathbf{s}) \approx \|\mathbf{s} - \widehat{\mathbf{s}_i}\|_2 \quad (4)$$

3.3 System Configuration

The experimental settings (unless otherwise stated) of the classification system were as follows: Thirteen MFCC's with the log-energy appended to MFCCs 1-12, The reduced MFCC set (MFCCr) was formed from MFCCs 1-5. The delta and delta-delta coefficients (Δ, ΔΔ) were extracted by the standard regression equation. Six PLP's, comprising the log-energy appended to PLP's 1-5, were employed. Eight LSP coefficients were computed from eight LPC coefficients. Twelve LPCC's and twenty SCF/SCA coefficients were extracted.

HTK was used to train the GMMs. For consistency of comparison across all features tested, 16 mixture components were used and 10 EM iterations with a variance flooring threshold set to 0.01. SVM modeling and testing was done using SVM*light*[3] and a linear kernel. For SRC, the solution of equation (2) was achieved using the GPSR[4] toolbox.

For the single-dimensional features SVM and SRC systems the input vectors were formed from the utterance based statistics of mean, max, min, standard deviation, median, first and third quartile, skewness and kurtosis. For the multi-dimensional SVM and SRC classifiers, GMM supervectors were used as the input to the classifiers.

3.3 Database

The database consists of 800 speakers and 18 utterance types discretised into a likeable class (L) or non-likable class (NL). For further details of either the ranking or partitioning of the database the reader is referred to [7].

4. Results

4.1 Baseline

The baseline results for the likelihood sub challenge are reproduced in Table 1.

Table 1. *Unweighted Accuracy Baseline for Challenge* [7]

System	ACC(UW)
SVM	58.5
Random Trees	57.6

4.2 Initial Testing

The results shown in Tables 2 and 3 are a subset taken from a wider range of tests using the openSMILE feature set. We have chosen to report these features as they were above chance level in one of the systems for either male or female class, with the exception of jitter, which is included for reference. Training data were used to form the background data used for the initial testing. In this phase the results were obtained using the development data set using 10 fold cross validation at a 70 - 30 training / testing split. All results reported are unweighted classification accuracy.

4.2.1 Single dimensional Features

For the GMM results, the average taken across the L and NL classes for both genders roughly follows what is seen in [1]. Spectral features were the best performing but we see voice quality features playing a more important role than prosodic. The results of shimmer match those reported in [6]. The higher accuracy of pitch direction could be explained through males and females using specific intonation and pitch patterns when trying to be liked [3, 6].

In the male classes, the poor NL F0 results could be explained through the results reported in [6]; males expand their pitch range when trying to be liked. We could hypothesize that NL male F0 could be a subset of the L, thus reducing class separation. We can also use [6] to help explain the evenness of the F0 and intensity results in females. Females potentially

2. *http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html*

3. *http://svmlight.joachims.org*

4. *http://www.lx.it.pt/~mtf/GPSR*

Table 2. *Likeability classification using single-dimension features*

FEATURES	GMM-UBM – (feature based)					SVM - (utterance based statistics)					SRC - (utterance based statistics)				
	Male		Female		*AVE*	Male		Female		*AVE*	Male		Female		*AVE*
	L	NL	L	NL		L	NL	L	NL		L	NL	L	NL	
f0	60.00	30.00	**50.00**	**51.54**	*47.88*	48.57	67.69	52.31	57.69	*56.57*	91.43	16.15	24.62	74.62	*51.70*
intensity	59.29	36.92	51.54	48.46	*49.05*	**62.14**	**50.77**	**53.08**	**63.08**	*57.27*	**67.86**	**52.31**	23.85	73.08	*54.28*
f0ds	**57.14**	**60.00**	**53.08**	**51.54**	*55.44*	60.71	46.15	**52.31**	**53.08**	*53.06*	72.86	24.62	39.23	80.00	*54.18*
Jitter	54.29	43.85	37.69	73.85	*52.42*	40.00	56.92	62.31	33.08	*48.08*	35.00	73.85	67.69	33.85	*52.60*
Shimmer	**56.43**	**63.85**	**59.23**	**75.38**	*63.72*	57.14	28.46	33.85	80.00	*49.86*	43.57	57.69	83.08	19.23	*50.89*
ZCR	57.14	33.08	64.52	40.00	*48.71*	59.29	40.00	59.23	40.77	*49.82*	79.29	21.54	80.00	28.46	*52.32*
SpectralRollOff25	**59.29**	**61.54**	**63.85**	**73.08**	*64.44*	49.29	60.00	67.69	34.62	*52.90*	35.00	62.31	64.62	40.77	*50.67*
Spectral Flux	52.86	40.77	**61.54**	**53.85**	*52.25*	40.00	62.31	**62.31**	**50.77**	*53.85*	68.57	40.00	**65.38**	**60.00**	*58.49*

encode likeability differently depending on the gender of the speaker they are talking to; this information is not available in the challenge database, so we are unable to test this hypothesis further.

The SVM system is the closest system we tested to that from which the baseline results were obtained [7], and here the prosodic features are the best performing. These results support the hypotheses in the previous paragraph.

The SRC results again roughly match those reported in [1], with spectral features performing best for both genders. Interestingly, prosodic features have better classification accuracy than voice quality. The results of the SRC are comparable with those of the SVM system, with the advantage that SRC requires no classifier training.

4.2.2 Multi-dimensional Features

For the GMM system, the poorer performance of the cepstral and detailed spectral features is not surprising given the results in [1], where cepstral features are the poorest performing class. The poorer female results are interesting as it is reported in [5] that likeable women have a more even energy spread over the spectrum.

The SVM results are somewhat surprising. The different configurations trialed all struggle to consistently break above chance for any feature compared across both classes for either gender. For the SRC, only the LPCC feature broke above chance across all classes in multi-dimensional systems.

For all three classifiers, we can see a difference in the performance between the two genders. This helps support the hypothesis that the two genders encode likeability differently.

4.3 Improvements to Multi-dimensional Systems

In general, MFCCs and LSPs are the most consistent features across not only different classification approaches but male and female subsets. This is not surprising as MFCCs represent their speaker's acoustic space with greater reliability than the other features tested [12]. LSPs are also a stable and robust feature especially when compared with the LP coefficients. Reduced MFCCs didn't perform as well, perhaps surprisingly given the spectral roll-off performance, although the importance of higher MFCC coefficients in classifying likeability is shown in [5].

There is a drop in performance when comparing the GMM classifier with both the SVM and SRC GMM supervector systems. This could be due the role that the covariances play when we model likeability; perhaps not enough of this information is preserved in the supervectors, or that the occupation counts are not very important and this is too heavily encoded in the supervectors. Hence, we tested MFCCs in both the SVM and SRC systems with different MAP relevance factor settings. The results of this are shown below in Table 4. Changing the relevance factor has minimal effect on the results for both systems; similar results have been reported for GMM speaker recognition systems [13].

Table 4. *Different MAP settings for MFCC in SVM and SRC*

Set Up	MALE		FEMALE		*AVE*
	L	NL	L	NL	
SVM	60.71	37.69	52.31	46.15	*49.22*
SVM 0.5xτ	65.71	31.54	64.62	46.15	*51.52*
SVM 2xτ	54.29	45.38	60.77	52.48	*50.73*
SRC	71.43	43.85	65.38	59.23	*59.97*
Sparse 0.5xτ	72.14	41.54	66.92	58.46	*59.77*
Sparse 2xτ	67.14	40.77	63.85	57.69	*57.36*

It has been shown in speaker recognition that better acoustic modeling can help improve system performance. In [12], a method is described for training a UBM by selecting a small amount of data, whilst still maintaining and even increasing system performance. We utilized this feature stability-based method to test whether UBM reduction provides a possible improvement. These tests were run for MFCCs, as they are a stable clustering feature, and the results, shown in Table 5, demonstrate that improvements in system performance can be obtained using a selective subset of training data for the UBM.

Table 5. *GMM likeability classification based on different UBM training data fractions*

% of optimized frames used	MALE		FEMALE		*AVE*
	L	NL	L	NL	
100	57.86	67.69	83.03	45.38	*63.50*
80	72.86	53.08	86.15	32.31	*61.10*
60	**74.29**	**56.15**	**82.31**	**53.85**	***66.65***
40	75.00	53.00	85.38	43.08	*64.33*
20	72.14	36.92	82.31	56.15	*61.88*

To further test the suitability of the given training data for use in a UBM we tested the challenge development set with two different UBMs; the first formed from the personality sub challenge data set [7] and the second formed from the NIST 2004 database. The overall system accuracy when using NIST 2004 UBM was comparable to that of our regular (non-optimized) UBM; this is not that surprising (apart from its American English language!) as this database is known to be suitable for use as background data. The results in both Tables 5 and 6 show the potential importance of optimizing UBM data

Table 3. *Likeability classification using multi-dimensional features*

FEATURES	GMM-UBM – (feature-based)					SVM – (supervector-based)					SRC – (supervector-based)				
	Male		Female		*AVE*	Male		Female		*AVE*	Male		Female		*AVE*
	L	NL	L	NL		L	NL	L	NL		L	NL	L	NL	
MFCC	**57.86**	**67.69**	83.03	45.38	*63.50*	60.71	37.69	52.31	46.15	*49.22*	71.43	43.85	65.38	59.23	*59.97*
MFCCr	**62.86**	**68.46**	59.23	43.85	*58.60*	67.86	32.31	50.00	49.23	*49.85*	67.86	48.46	40.00	40.77	*49.27*
PLP	45.00	49.23	36.92	49.23	*45.10*	73.57	40.77	52.31	47.69	*53.59*	**69.29**	**56.15**	44.62	41.54	*52.90*
LSP	**52.14**	**59.23**	56.92	49.23	*54.38*	60.00	47.69	44.62	52.31	*51.15*	78.57	21.54	**63.08**	**52.31**	*53.87*
LPCC	**75.00**	**56.92**	72.31	28.46	*58.17*	75.00	33.82	64.62	46.92	*55.10*	**58.57**	**50.00**	**55.38**	**55.38**	*54.84*
SCF	**60.00**	**69.23**	73.85	45.38	*62.12*	62.14	35.38	59.23	48.46	*51.30*	98.23	5.88	78.46	25.38	*51.99*
SCA	**52.86**	**55.38**	59.23	37.69	*51.29*	80.71	29.23	**61.54**	**63.85**	*58.83*	53.57	49.23	**63.08**	**53.85**	*54.93*

selection in paralinguistic classification, and whilst the overall average of the different systems doesn't change greatly, differences in system performance can be seen when comparing accuracies on a per-gender / per-class basis.

Table 6. *GMM likeability classification using different UBMs*

UBM DATA	MALE		FEMALE		*AVE*
	L	NL	L	NL	
Likeability UBM (100%)	57.86	67.69	83.03	45.38	*63.50*
Personality UBM	72.86	56.15	83.08	46.92	*63.50*
NIST 2004 UBM	55.00	67.84	67.69	63.85	*63.60*

As reported in [3, 6], males and females use different intonation and pitch patterns when trying to be liked. Dynamic information could therefore be a useful addition to our multi-dimensional classifier; hence we tested the GMM MFCC system using the 60% optimized UBM with Δ and $\Delta\Delta$ coefficients, extracting the Δ and $\Delta\Delta$ coefficients using different window lengths ($2K-1$). Our testing showed mixed results for both genders with an increase in *NL* classification but a slight decrease in *L* classification. We were able to increase male and female NL detection an average of 7% above our baseline (in relative terms) with window sizes of $K = 4$ and $K = 9$.

Our final test was whether normalization could be beneficial. We tested the GMM MFCC system using the 60% optimized UBM and tested mean, mean-variance and CDM normalization. Table 7 shows that normalization is not a useful addition. We can draw similar conclusions to the role of speaker normalization to those seen in [8].

Table 7. *Feature Warping on MFCC's*

Set Up	MALE		FEMALE		*AVE*
	L	N	L	N	
Un-normalized	74.29	55.38	75.38	50.77	*63.96*
MEAN	70.71	36.15	96.15	30.77	*58.45*
MeanVar	64.29	54.62	90.00	42.31	*62.80*
CDM	76.43	40.77	84.62	53.85	*63.91*

5. Challenge Systems and Results

For our final entry to the challenge, we chose five different systems. The first (*Sys 1*) is composed by fusing the best performing single dimensional features; F0dr, shimmer and SpectralRollOff25 using a GMM back-end adapted from the non-optimized UBM. The second system (*Sys 2*) was composed of MFCCs using a GMM back-end adapted from the 60% optimized UBM and includes delta coefficients extracted with $K = 9$ for males but no delta coefficients for females. The results of this testing and on the fusion of the two systems (*Sys 3*) are shown in Table 8.

Table 8. *GMM likeability classification on dev. and test sets*

System	Develop ACC(UW)	TEST ACC(UW)
1.	65.00	50.123
2.	68.27	51.800
3.	68.85	48.061

To test the performance of SRC we chose two system setups; firstly (*Sys 4*) employing MFCC supervectors adapted from a non-optimized UBM, and secondly (*Sys 5*) MFCC supervectors adapted from the 60% optimized UBM. We chose to include only highly scored likeable and not likeable speakers from development set in both dictionaries (top 20 highly scored male and female in both classes; 40 speakers per dictionary). Note that the development results in Table 9 are not cross-validated.

Table 9. *SRC likeability classification on dev. and test sets*

System	Develop ACC(UW)	TEST ACC(UW)
4	64.60	54.52
5.	64.98	52.19

Interestingly the best classifier on the test set was *Sys* 4. Given the abstract nature of the classification task, this result might be explained through the more robust nature of the sparse classification [10].

6. Conclusion

Classifying likeability proved to be a very difficult challenge, seen clearly in the lack of generalization of results from development set to test set. This could be due in part to the fact that likeability and its perception is a more abstract and personalized speaker trait, reflected in the varying cross correlation scores from the listeners who ranked the likeability database [7]. The single dimensional features of F0dr, Shimmer and SpectralRollOff25 were suited to the classification task for the development set but we were unable to reproduce this performance in the final classification. Whilst it is possible that we tuned our classifier so that it gained optimal classification when tested against the development set, we were careful to avoid potential system overtraining by using MAP adaptation and training our GMM's with a small number of parameters. Potentially, if we repeated the initial testing with the test set included, we could find another set of discriminating features. Given the benefits gained when testing on the development data, UBM optimization is worth exploring further as potential aid to paralinguistic classification. The 60% optimized UBM worked well with the development set with but not for the test set. SRC shows potential as a paralinguistic classifier: it gave our best system results on the test set and its performance on the development set was close to that of SVM. Future work will include testing both of these methodologies on other paralinguistic traits to further assess their suitability in this field.

7. References

[1] Burkhardt, F., et al., "Would You Buy A Car From Me?" – On the Likability of Telephone Voices", *InterSpeech 2011*, 2011: p. 1557 - 1560.

[2] Mairesse, F., et al., "Using linguistic cues for the automatic recognition of personality in conversation and text", *Journal of Artificial Intelligence Research*, 2007. 30: p. 457-500.

[3] Jurafsky, D., R. Ranganath, and D. McFarland, "Extracting social meaning: identifying interactional style in spoken conversation", *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2009: p. 638-646.

[4] Pinto-Coelho, L., et al., "On the development of an automatic voice pleasantness classification and intensity estimation system", *Computer Speech & Language*, 2012: p. not available.

[5] Weiss, B. and F. Burkhardt, "Voice attributes affecting likability perception", *InterSpeech 2010*, 2010: p. 2014-2017.

[6] Gravano, A., et al., "Acoustic and prosodic correlates of social behavior", *InterSpeech 2011,* 2011: p. 97-100.

[7] Schuller, B., et al., "The INTERSPEECH 2012 Speaker Trait Challenge", *InterSpeech 2012*, 2012: p. not available.

[8] Cummins, N., et al., "An Investigation of Depressed Speech Detection: Features and Normalization", *InterSpeech 2011*, 2011: p. 2997 - 3000.

[9] Kua, J.M.K., et al. "Speaker verification using sparse representation classification", *in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. 2011: p. not available.

[10] Wright, J., et al., "Robust Face Recognition via Sparse Representation", *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 2009. 31(2): p. 210-227.

[11] Donoho, D.L., "Compressed sensing", *IEEE Transactions on Information Theory*, 2006. 52(4): p. 1289-1306.

[12] Kua, J.M.K., et al. "Using clustering comparison measures for speaker recognition", *in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on.* 2011: p. not available.

[13] Reynolds, D.A., T.F. Quatieri, and R.B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models", *Digital Signal Processing*, 2000. 10(1-3): p. 19-41.