# Evaluating a Forensic Speaker Recognition using MFCC and LPC-residual

*Jose B. Trangol C.*[1], *Abel Herrera C.*[2]
[1]*Department of signal processing, Engineering School UNAM, Mexico*
jtrangol@yahoo.com
[2]*Department of signal processing, Engineering School UNAM, Mexico*
abelhc@hotmail.com

In forensic speaker recognition, different methods can be applied to determine if the unknown voice of the questioned recording (trace) belongs to the suspected speaker (source) (A. Drygajlo et al. 2003). Forensic speaker recognition shows very good performance in discriminating between voices of speakers under controlled recording conditions and sufficient data. However, the conditions in which recordings are made in investigative activities (e.g., anonymous calls and wire-tapping) cannot be controlled and pose a challenge to automatic speaker recognition (A. Alexander. 2005).

In this work we compare Mel-Frequency Cepstrum Coefficients (MFCC) and y Linear Predictive Coding-Residual (LPC-residual), MFCC are one of the most popular in speech processing (Md. R. Hasan et al. 2004). In the case of statistical modelling, the speech speaker can be represented by a Gaussian mixture model (GMM). The speaker recognition chosen for the following experiment is based on a text-independent. The first aim of the experiment was to test LPC residual and MFCC in a forensic speaker recognition task, was used short time of recording in questioned recording (trace) and suspected recordings (source), the second aim is to choose the minimum number of GMM component necessary to adequately model a speaker (Reynolds et al. 1995), 64 are appropriate for this base data to have a better performance in speaker discrimination. To be realistic, probably in forensic discriminations the most important is to allow for comparison of non-contemporaneous speech samples. Offender and suspect speech samples are usually separated by more a greater time span than is encompassed in a single recording session (Phil Rose et al. 2006). In this work 23 Mexican male with an age range of 18 to 29, all native speakers of Mexican Spanish, were recorded in two non-contemporaneous recordings, each speaker was recorded on two separate occasions separated by approximately one month, in order to attempt a forensically realistic discrimination. Since LPC residual has encompasses a large amount of data, a reduction of data was performed, the mean and variance was applied to reduce the number of data.

With 23 speakers, there were 23 non-contemporaneous same-speaker comparisons. 23 speakers gives make 506 different-speaker comparisons. 10, 15 and 30 second of training time was used and 10, 20 and 35 second of test time was used in the speaker recognition task. Likelihood ratio-based discriminations were then carried out. The identified speaker of each segment was compared to the actual speaker of the utterance and the number of segments which were correctly identified was tabulated. (Reynolds et al. 1995).

**Table 1.** We can see the correct percentage in the recognition of speakers, using MFCC and LPC-residual, using different time in training and testing.

| *MFCC* | *Testing* | *Testing* | *Testing* | *LPC-resid* | *Testing* | *Testing* | *Testing* |
|---|---|---|---|---|---|---|---|
| Training | 10 s | 20 s | 35 s | Training | 10 s | 20 s | 35 s |
| 10 s | 56.52 | 65.22 | 69.57 | 10 s | 60.87 | 65.22 | 73.91 |
| 15 s | 60.87 | 69.57 | 73.91 | 15 s | 65.22 | 73.91 | 78.26 |
| 30 s | 78.26 | 82.61 | 91.30 | 30 s | 69.57 | 78.26 | 86.96 |

The first aim of the experiment was to test LPC residual in the task of forensic speaker recognition, LPC-residual presents a discrimination significant with different speakers, for different time recordings, but more experiments must be performed. A second conclusion is to combine LPC and LPC-residual in a future work.

## References


A. Drygajlo, D. Meuwly, A. Alexander, (2003). Statistical Methods and Bayesian Interpretation of Evidence in Forensic Automatic Speaker Recognition, in Proceedings of Eurospeech 2003, Geneva, Switzerland, pp. 689–692.

A. Alexander, (2005). Forensic Automatic Speaker Recognition using Bayesian Interpretation and Statistical Compensation for Mismatched Conditions, Ph.D. thesis, Swiss Federal Institute of Technology at Lausanne.

P. Rose, Y. Kinoshita, T. Aldemar. (2006). Realistic Extrinsic Forensic Speaker Discrimination with the Diphthong /ar/. Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. University of Auckland, New Zealand. December 6-8. Copyright, Australian Speech Science & Technology Association Inc.

Md. R. Hasan, M. Jamil, Md. Golam Rabbani Md. Saifur Rahman, (2004). Speaker identification using mel frequency cepstral coefficients, 3rd International Conference on Electrical & Computer Engineering, CECE.

Reynolds, D. A. and Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models, IEEE Trans. Speech Audio Process. 3, 72–83