Identities, forgeries and disguises

Gérard Chollet, Hervé Bredin, Sanjay Kanade, Walid Karam,
Chafic Mokbel, Patrick Perrot, and Dijana Petrovska-Delacrétaz

gerard.chollet@telecom-paristech.fr,herve.bredin@irit.fr,
sanjay.kanade@it-sudparis.eu,walid.karam@balamand.edu.lb,
chafic.mokbel@balamand.edu.lb,perrot@telecom-paristech.fr,
dijana.petrovska@it-sudparis.eu

1 Introduction

The preservation of your identity is a fundamental right. In many situations, you need to claim your identity and this claim needs to be verified somehow. The technology called biometrics may help. But, what if a deliberate impostor claims your identity? Will this forgery be always detected? Biometric identity verification is imperfect. This presentation will review some of the techniques that a deliberate impostor could use to defeat a biometric verification system. It will focuss on audio-visual forgeries using voice conversion and face animation. Such techniques also find useful applications in multimedia.
In a second part, we'll discuss the techniques which could be used if somebody wants to hide his/her identity. Again, voice conversion and face animation is quite efficient to disguise your identity. We'll review some of the techniques to detect voice disguise which could be used in forensic applications.
Finaly, we'll take a look at 'cancellable biometrics', a set of techniques to allow users renew their biometric templates and models in case these have been stolen and used by impostors.

2 Audio-visual forgery

The identity of a person is primarily determined visually by his face and audibly by his voice. These two modalities, i.e. face and voice, are used naturally by people to recognize each other. They are also employed by many identity recognition systems to automatically verify or identify humans for commercial, security and legal applications, including forensics. Audio-visual identity verification is introduced in [2]. However, altering the features of the face and/or the voice can be effectively used to trick an audio-visual identity verification system so as to have an impersonator be accepted as a genuine user.

Audio-visual forgery, or imposture, is the process of modifying both the face and the voice of an impostor to make them look and sound like those of an authentic client. It is reasonable to assume that an impostor has knowledge of the audio-visual recognition system techniques used on one hand, and, on the other hand, has enough information about the target client (face image, video

sequence, voice recording.) It would then be possible to use techniques of voice transformation and face animation as deliberate imposture methods to defeat the audio-visual system. Karam et al. [20, 21] propose voice transformation and face animation techniques to evaluate the effects of deliberate imposture on identity verification systems.

At the audio level, a voice transformation technique (MixTrans) [20] is employed to change the perceived speaker identity of the speech signal of the impostor to that of the target client. At the visual level, an animation of the impostor face is achieved using a thin-plate spline warping [20]. Face animation is equally achieved using commercial animation packages such as CrazyTalk. Abboud et al. [1] propose appearance-based lip tracking and cloning on speaking faces as a means of face transformation. Imposture results indicate an increase in equal error rates from 5.1% to 15.39% on the performance of the audio-visual verification system, implying a higher impostor acceptance rate.

To overcome the challenges imposed by deliberate imposture on audio-visual identity verification systems, Bredin et al. in [4, 8, 5, 6] provide a study on the level of audio visual synchronization as a means of imposture detection, which helps make talking face authentication robust to deliberate imposture.

Audio-visual identity verification systems and effects of deliberate imposture were reported within the framework of the Biosecure project [9, 7, 11]. A guide to biometric reference systems and performance evaluations is provided in [29].

3 Identity disguise

Voice disguise is considered as a deliberated action of the speaker who wants to falsify or to conceal his identity [28, 27]. A relevant way to mask his identity is to use a simple but efficient disguise. Lots of possibilities are offered to a speaker to change his voice. In the field of automatic speaker recognition application, one of the most threats is voice disguise. Based on crime analysis in Germany, it is noted in [23] that there was "... an overall occurrence of voice disguise in 52 percent of the cases where the offender used his/her voice and may have expected to have it recorded during the criminal action. This percentage includes cases of blackmailing, where the specific percentage was as high as 69 percent." And in [10] it is noted that "Disguised speech is typically found in situations in which the criminal thinks he is being recorded. This situation is very common in cases of kidnapping, a kind of crime whose incidence has increased considerably in the past years in Brazil."

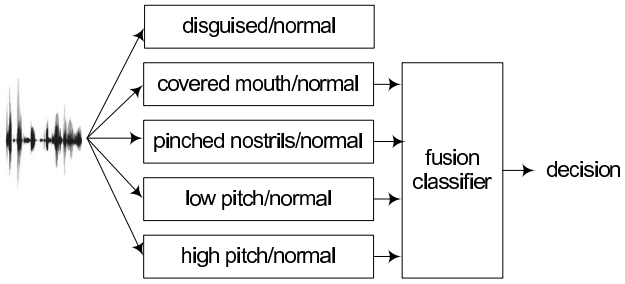
This section focuses on four specific non electronic and deliberated disguises according to the classification proposed by [32]: High pitched voice, low pitched voice, covered mouth, and pinched nostrils [27]. An experiment based on a 50 speaker database is proposed. A set of acoustic features in the speech signal, including the formants F1 and F2, mean F0, min F0, max F0, 12 MFCC and their first derivatives are extracted from each speech segment. Different classification techniques are evaluated to detect disguise: k-nearest neighbours, Support Vector Machine (SVM) and a two different architectures of classifier fusion. The first

architecture consists in a parallel combination of five classifiers as proposed in Fig. 1(a).

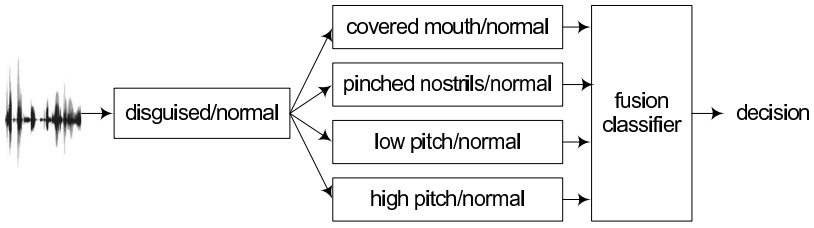
The second architecture consists in a hybrid combination of classifiers as proposed in Fig. 1(b). The level of performance of each classifier is based on the analysis of ROC (Receiver Operating Characteristic) curve and the criterion linked to the curve, the Area Under Curve (AUC). The ROC graph is a useful technique for organizing classifiers and visualizing their performance.

Fig. 2 proposes the results of each classifier between normal voice and a combined of four disguises composed by the different disguises previously described. This curve reveals a good level of performance for the parallel architecture and the SVM classifier with an AUC of 0.79 and 0.78. The hybrid architecture presents an AUC of 0.72 and the 5 nearest neighbours an AUC of 0.67. In the area of forensic speaker recognition it could be interesting to realize a step of disguise detection as pre-processing in order to avoid directing the investigation toward unlikely suspects and away from likely ones.

A more sophisticated method to disguise his voice is to use voice conversion considered as an electronic-deliberated technique. In [23] a proposition of original voice conversion technique to trick an automatic speaker recognition system is proposed where a degradation of near than 50% is noticed on the level of performance after the conversion. In [26] three different methods of voice conversion are compared in face of a forensic application: imitation of a French politician.



(a) Parallel classifier fusion



(b) Hybrid classifier fusion

Fig. 1. Classification architectures

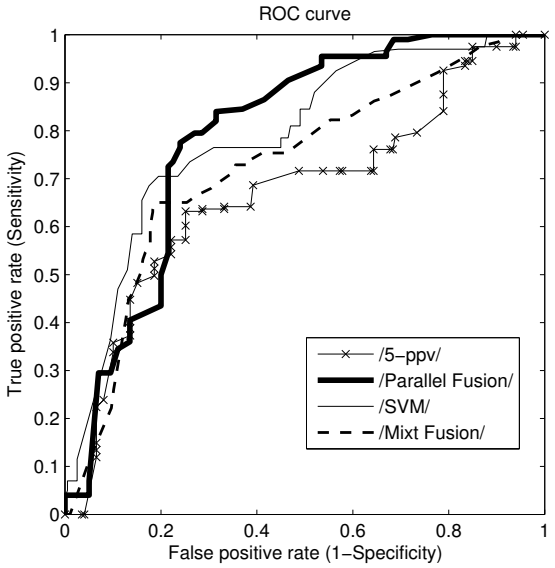


Fig. 2. Performance of normal and disguised voices

4 Cancelable biometrics

The biometric characteristics are permanently associated with the user. Therefore, if a biometric characteristic is compromised, the stored template cannot be revoked meaning that it cannot be replaced with a new one. Thus biometric systems lack revocability. Another problem related with the permanence of biometric data is the possibility of cross-database matching. In classical biometric systems, the information needed for further comparisons, denoted as biometric reference or template, is stored in a database. This information remains substantially similar across databases if the modality and the biometric algorithm are the same (e.g., fingerprint minutiae set extracted from the same fingerprint in different systems are similar). Therefore, a compromised template from one biometric database can be used to access information from another system which uses the same biometric modality. This can be considered as a threat to privacy. Moreover, in some cases, the stored information can be used to create a dummy representation of the biometric trait which can be used to access the system.

Many solutions, termed as cancelable biometrics, are proposed in order to overcome the problems of revocability and cross-database matching of biometrics [30, 31, 13, 22, 33, 3]. These solutions basically involve combination of some assigned user specific secret with the biometric characteristics. The use of an assigned secret allows the revocation of the template if the template is compromised.

Kanade et al. [16, 17] proposed a simple shuffling scheme which randomizes the biometric data with the help of a shuffling key. The data to be shuffled is divided into blocks and these blocks are rearranged according to the bit values of the shuffling key. The advantages of this scheme are: (a) induces revocability in biometric systems, (b) improves the verification performance (nearly 80% decrease in equal error rate) because it increases the impostor Hamming distance without changing the genuine Hamming distance, (c) template diversity, (d) makes cross-matching impossible and thus protects privacy.

There are also a number of systems in literature with which a long and stable bit-string can be derived from biometrics [15, 14, 12, 24]. Such systems can also possess the properties of revocability, template diversity, and privacy protection.

Kanade et al. proposed such key regeneration systems using uni-biometrics (iris [16]) as well as multi-biometrics (two iris system [18] and multi-modal system using a combination of iris and face [19]).

Such systems can help if an impostor has stolen the biometric data. In this case, these systems allow revocation of the compromised template and re-enrollment of the user with the same biometric data which is not possible with classical biometric systems.

Keywords: Identity verification, forgery, audio-visual imposture, voice conversion, face animation, voice disguise, cancelable biometrics

References

1. B. Abboud and G. Chollet, "Appearance based lip tracking and cloning on speaking faces," in *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, 15-17 2005, pp. 301 – 305.
2. B. Abboud, H. Bredin, G. Aversano, and G. Chollet, "Audio-visual identity verification: an introductory overview," in *Progress in nonlinear speech processing*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 118–134.
3. T. E. Boult, W. J. Scheirer, and R. Woodworth, "Revocable fingerprint biotokens: Accuracy and security analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.
4. H. Bredin and G. Chollet, "Measuring audio and visual speech synchrony: Methods and applications," in *Visual Information Engineering, 2006. VIE 2006. IET International Conference on*, 26-28 2006, pp. 255 –260.
5. ——, "Audiovisual speech synchrony measure: application to biometrics," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 179–179, 2007.
6. ——, "Making talking-face authentication robust to deliberate imposture," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, vol. 1, pp. 1693–1696, 31 2008-April 4 2008.
7. H. Bredin, G. Aversano, C. Mokbel, and G. Chollet, "The biosecure talking-face reference system," in *In 2nd Workshop on Multimodal User Authentication*, 2006.
8. H. Bredin, A. Miguel, I. H. Witten, and G. Chollet, "Detecting replay attacks in audiovisual identity verification," in *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2006, pp. 621–624.

9. G. Chollet, G. Aversano, B. Dorizzi, and D. Petrovska-Delacrétaz, "The first biosecure residential workshop," in *4th International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, September 2005.
10. R. de Figueiredo and H. de Souza Britto, "A report on the acoustic effects of one type of disguise," *Forensic Linguistics*, vol. 3, no. 1, pp. 168–175, 1996.
11. B. Fauve, H. Bredin, W. Karam, F. Verdet, A. Mayoue, G. Chollet, J.Hennebert, R.Lewis, J. Mason, C. Mokbel, and D. Petrovska, "Some results from the biosecure talking face evaluation campaign," in *International Conference on Accoustics, Speech, and Signal Processing, ICASSP*. IEEE, 2008, pp. 4137–4140.
12. F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081–1088, 2006.
13. A. T. B. Jin, D. Ngo, C. Ling, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition*, vol. 37, no. 11, pp. 2245–2255, November 2004.
14. A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. IEEE Int. Symp. Information Theory*, A. Lapidoth and E. Teletar, Eds. IEEE Press, 2002, p. 408.
15. A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proceedings of the Sixth ACM Conference on Computer and communication Security (CCCS)*, 1999, pp. 28–36.
16. S. Kanade, D. Camara, E. Krichen, D. Petrovska-Delacrétaz, and B. Dorizzi, "Three Factor Scheme for Biometric-Based Cryptographic Key Regeneration Using Iris," in *The 6th Biometrics Symposium (BSYM)*, September 2008.
17. S. Kanade, D. Petrovska-Delacrétaz, and B. Dorizzi, "Cancelable Iris Biometrics and Using Error Correcting Codes to Reduce Variability in Biometric Data," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2009.
18. ——, "Multi-Biometrics Based Cryptographic Key Regeneration Scheme," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, September 2009.
19. ——, "Obtaining Cryptographic Keys Using Feature Level Fusion of Iris and Face Biometrics for Secure User Authentication," in *IEEE CVPR Workshop on Biometrics*, June 2010.
20. W. Karam, H. Bredin, H. Greige, G. Chollet, and C. Mokbel, "Talking-face identity verification, audiovisual forgery, and robustness issues," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 746481, p. 18, 2009.
21. W. Karam, C. Mokbel, H. Greige, and G. Chollet, "Audio-visual identity verification and robustness to imposture," in *Advances in Biometrics, Third International Conference, ICB 2009, Alghero, Italy*, ser. Lecture Notes in Computer Science, M. Tistarelli and M. S. Nixon, Eds., vol. 5558. Springer, June 2009, pp. 796–805.
22. A. Lumini and L. Nanni, "An improved biohashing for human authentication," *Pattern Recognition*, vol. 40, no. 3, pp. 1057–1065, March 2007.
23. H. Masthoff, "A report on a voice disguise experiment," *Forensic Linguistics*, vol. 3, no. 1, pp. 160–167, 1996.
24. K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance," *IEEE Transactions of Information Forensics and Security*, vol. 2, no. 4, pp. 744–757, December 2007.
25. P. Perrot, G. Aversano, R. Blouet, M. Charbit, and G. Chollet, "Voice forgery using alisp: Indexation in a client memory," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05*, vol. 1, 2005, pp. 17–20.

26. P. Perrot, M. Morel, J. Razik, and G. Chollet, "Vocal forgery in forensic sciences," in *Forensics in Telecommunications, Information and Multimedia, Second International Conference, e-Forensics 2009, Adelaide, Australia, Revised Selected Papers*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, M. Sorell, Ed., vol. 8. Springer, January 2009, pp. 179–185.
27. P. Perrot and G. Chollet, "The question of disguised voices," in *Acoustics08*, July 2008.
28. P. Perrot, C. Preteux, S. Vasseur, and G. Chollet, "Detection and recognition of voice disguise," in *proceedings of International Association for Forensic Phonetics and Acoustics Conference 2007*, Plymouth, UK, July 2007, p. 3.
29. D. Petrovska-Delacrétaz, G. Chollet, and B. Dorizzi, *Guide to Biometric Reference Systems and Performance Evaluation*. Springer Publishing Company, Incorporated, 2009.
30. N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.
31. N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating cancelable fingerprint templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 561–572, April 2007.
32. R. D. Rodman, "Speaker recognition of disguised voices," in *Consortium on Speech Technology Conference on Speaker Recognition by Man and Machine: Directions for Forensic Applications COST250*, 1998.
33. M. Savvides, B. V. Kumar, and P. Khosla, "Cancelable biometric filters for face recognition," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR04)*, vol. 3, August 2004, pp. 922–925.