A new speaker verification spoofing countermeasure based on local binary patterns

Federico Alegre, Ravichander Vipperla, Asmaa Amehraye, Nicholas Evans

A new speaker verification spoofing countermeasure based on local binary patterns

Federico Alegre[1], *Ravichander Vipperla*[2†], *Asmaa Amehraye*[1] *and Nicholas Evans*[1]

[1]Multimedia Communications Department, EURECOM, Sophia Antipolis, France
[2]Nuance Communications, Cambridge, UK

{alegre,fillatre,evans}@eurecom.fr, ravichander.vipperla@nuance.com

Abstract

This paper presents a new countermeasure for the protection of automatic speaker verification systems from spoofed, converted voice signals. The new countermeasure is based on the analysis of a sequence of acoustic feature vectors using Local Binary Patterns (LBPs). Compared to existing approaches the new countermeasure is less reliant on prior knowledge and affords robust protection from not only voice conversion, for which it is optimised, but also spoofing attacks from speech synthesis and artificial signals, all of which otherwise provoke significant increases in false acceptance. The work highlights the difficulty in detecting converted voice and also discusses the need for formal evaluations to develop new countermeasures which are less reliant on prior knowledge and thus more reflective of practical use cases.

Index Terms: speaker verification, biometrics, imposture, countermeasures, local binary patterns

1. Introduction

Text-independent, automatic speaker verification (ASV) systems are widely acknowledged to be vulnerable to spoofing. Previous work over the last decade has considered classical attacks such as impersonation [1, 2] and replay [3, 4], in addition to more sophisticated attacks involving speech synthesis [5, 6], voice conversion [7–10] and artificial signals [11]. All provoke significant increases in the false acceptance rate of state-of-the-art ASV systems. It is only relatively recently that the community has investigated spoofing countermeasures, as has been the case for other biometric modalities, e.g. face recognition [12, 13].

Characteristic to almost all previous work specific to ASV is the assumption of prior knowledge, i.e. the nature of the attack is assumed to be known. This assumption is unrealistic; in practice the spoofing attack can never be known and then the performance of existing countermeasures in practical scenarios cannot be guaranteed. As an example we consider previous work based on the use of phase [14–16] and prosodic features [17, 18] as a means of detecting voice conversion and speech synthesis attacks. The particular approach to voice conversion investigated in [9] essentially modifies only the spectral slope of a converted utterance while retaining the phase and pitch of the original, genuine speech signal. As such, it will likely overcome the countermeasures proposed in [14–18]. Spoofing thus remains very much an open problem.

This paper presents a new countermeasure which aims to provide a more universal spoofing countermeasure which is less dependent on prior knowledge, i.e. not specific to a given attack. It is based on characteristics of a sequence of feature vectors captured using Local Binary Patterns (LBP) [19], a popular approach to texture analysis in image processing and especially face recognition [20]. While the approach was optimised for the detection of converted voice, it is also effective in detecting synthesized speech and artificial signals. Compared to a previously reported spoofing countermeasure [21] the new LBP-based countermeasure is shown to give significantly better performance across three different spoofing attacks and is thus considered to be more generalised than previous solutions. It operates on conventional acoustic features, is computationally efficient and readily integrated into any standard ASV system.

The remainder of this paper is organized as follows. Spoofing attacks and the new countermeasure are presented in Sections 2 and 3, respectively. Experimental work is described in Section 4. Our conclusions are presented in Section 5.

2. Spoofing attacks

In this section we describe our approach to voice conversion, speech synthesis and attacks with artificial signals.

2.1. Voice Conversion

All work involving voice conversion was performed with our own implementation of the approach originally proposed in [9]. It was developed to test the limits of ASV when the vocal tract information in the speech signal of a spoofer is converted towards that of another, target person. At the frame level, the speech signal of a spoofer denoted by $y(t)$ is filtered in the spectral domain as follows:

$$Y'(f) = \frac{|H_x(f)|}{|H_y(f)|} Y(f) \quad (1)$$

where $H_x(f)$ and $H_y(f)$ are the vocal tract transfer functions of the targeted speaker and the spoofer respectively. $Y(f)$ is the spoofer's speech signal whereas $Y'(f)$ denotes the result after voice conversion. As such, $y(t)$ is mapped or converted towards the target speaker in a spectral-slope sense. As we show later, this is sufficient to overcome most ASV systems.

$H_x(f)$ is determined from a set of two Gaussian mixture models (GMMs). The first, denoted as the automatic speaker recognition (asr) model in the original work, is related to ASV feature space and utilized for the calculation of a posteriori probabilities whereas the second, denoted as the filtering (fil)

[†]This author's contribution to the work was made while employed at EURECOM

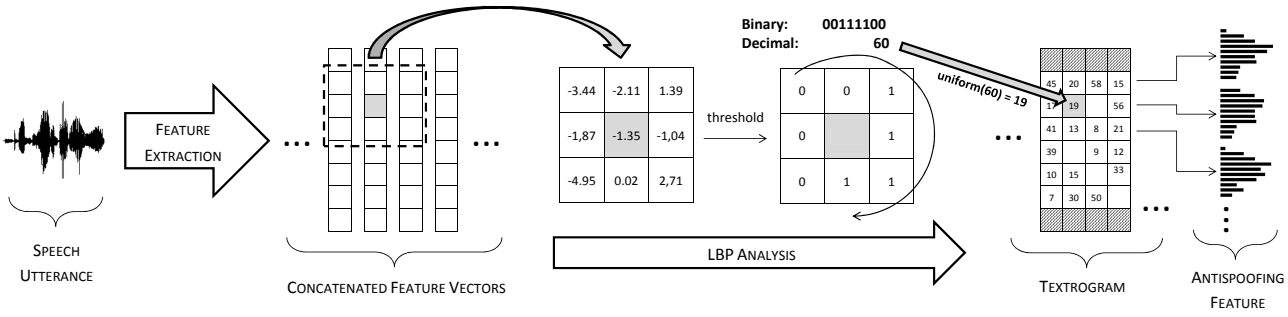

Figure 1: Application of uniform LBP analysis to obtain a textrogram from a matrix formed from the concatenation of conventional feature vectors. Non-uniform patterns (blank cells in textrogram) are discarded and the resulting feature used for spoofing detection is formed from the concatenation of normalised histograms of the remaining uniform codes in each row.

model, is a tied model of linear predictive cepstral coding (LPCC) coefficients from which $H_x(f)$ is derived. LPCC filter parameters are obtained according to:

$$x_{fil} = \sum_{i=1}^{M} p(g^i_{asr}|y_{asr})\mu^i_{fil} \quad (2)$$

where $p(g^i_{asr}|y_{asr})$ is the a posteriori probability of Gaussian component g^i_{asr} given the frame y_{asr} and μ^i_{fil} is the mean of component g^i_{fil} which is tied to g^i_{asr}. $H_x(f)$ is estimated from x_{fil} using an LPCC-to-LPC transformation and a time-domain signal is synthesized from converted frames with a standard overlap-add technique. Full details can be found in [9, 22, 23].

2.2. Speech synthesis

One of the main advantages of statistical parametric speech synthesis using hidden Markov models (HMMs) is its ability to adapt to a target voice with little adaptation data. This hence becomes a powerful tool for generating spoofing attacks. In this paper, we use a parametric speech synthesis system as described in [24] where the spectral and excitation parameters of speech are simultaneously modeled along with explicit modeling of duration probabilities using multi-space distribution hidden semi Markov models (MSD-HSMM). The speaker independent model parameters and excitation are adapted to the target speaker via the constrained structural maximum a posteriori linear regression (CSMAPLR) [25] method. The spoofing speech signals are synthesized using a vocoder based on the STRAIGHT method [26] using the target speaker adapted MSD-HSMMs and multiband excitation.

2.3. Artificial signals

Artificial signal attacks are based on the algorithm reported in [11]. It is based on a modification of the voice conversion algorithm presented in Section 2.1.

Let $S = \{c_1, ..., c_n\}$ be a short sequence of consecutive speech frames selected from an utterance of the targeted speaker. The algorithm seeks a new sequence of speech frames S^* which maximises the score of a given ASV system and thus the potential for spoofing. Here and below the '*' symbol indicates an optimised quantity.

Each frame $c(t)$ belonging to S is initially transformed in the frequency domain with voice conversion where we now have:

$$C'(f) = \frac{|H^*_c(f)|}{|H_c(f)|} C(f) \quad (3)$$

Optimisation is then applied to identify a set of filters $H^*_S = \{H^*_{c_1}(f), H^*_{c_2}(f), ..., H^*_{c_n}(f)\}$. Instead of estimating each filter independently using Equation 2, however, the set of filters is jointly optimized using a genetic algorithm. Full details are presented in [11].

3. Spoofing countermeasure

The new countermeasure proposed in this paper was designed with prior knowledge of a specific spoofing attack, in this case, voice conversion. We acknowledge that such a setup is not representative of the practical use case (where the exact nature of the spoofing attack can never be known) but note this to be the case with all previous work; there are currently no standard datasets for spoofing and countermeasure assessment. Also new to this paper is a further analysis of countermeasure performance on alternative, 'unseen' spoofing attacks, for which the countermeasure was not optimised. While we have worked with such signals previously, and in this sense they are not truly 'unseen,' the experiments with speech synthesis and artificial signal attacks give some insight into the potential of more generalised countermeasure solutions.

The new countermeasure is based on the hypothesis that modifications made through spoofing disturb the natural, dynamic spectro-temporal 'texture' of genuine speech. Motivated by the fact that computer vision techniques were already successfully applied in the speech field [27], we have investigated the application of a standard texture analysis approach, known as Local Binary Patterns [19], to a 2-dimensional 'image' of a speech utterance, where here the image is a linear-scaled cepstrogram appended with dynamic features.

The standard Local Binary Pattern (LBP) operator [19] is a non-parametric, 3x3 kernel which assigns a binary code to each pixel in an image according to the comparison of its intensity value to that of its eight surrounding pixels. The procedure is illustrated in Figure 1. A binary value of '1' is assigned when the intensity of neighbouring pixels (here feature components) is higher, whereas a value of '0' is assigned when neighbouring pixels are of lower or equal intensity. Each pixel is thus assigned one of $2^8 = 256$ binary patterns.

In this work we reduce the number of possible patterns

according to the standard Uniform LBP approach described in [19]. Uniform LBPs are the subset of 58 patterns which contain at most two bitwise transitions from 0 to 1 or 1 to 0 when the bit pattern is traversed in circular fashion. As an example, the subset includes patterns 00000001 and 00111100 but not 00110001. As reported by [19], most patterns are naturally uniform and empirical evidence suggests that their use in many image recognition applications leads to better performance than the full set of uniform and non-uniform patterns. We observed similar findings in our work and thus pixels corresponding to any of the 198 non-uniform patterns are simply ignored.

LBPs are determined for each pixel in the linear-scaled cepstrogram thus resulting in a new matrix of reduced dynamic range, here referred to as a 'textrogram'. The textrogram captures short-time feature motion beyond that in conventional dynamic parametrizations. The LBP-based countermeasure is based on concatenated histograms formed from the pixel values across each row in the textrogram. The histograms are individually normalised and their resulting bin values are stacked vertically to obtain a new vector in the same manner as GMM mean-vectors are stacked to form supervectors. The division of the textrogram (or equivalent in image recognition problems) is also standard practice [20] and serves to provide a greater level of granularity than would be provided with only a single histogram corresponding to the full textrogram.

The countermeasure is integrated into a full ASV system as an independent classifier in equivalent fashion to the work in [6, 15, 21]. LBP-based features are calculated for the test data and that used for training client models. The two resulting feature vectors are compared using histogram intersection and the resulting score is thresholded to classify the test signal as genuine speech or a spoofing attack.

4. Experimental work

Here we report experimental work which assesses the performance of the new LBP-based countermeasure. Results are compared with those similarly obtained using a pair-wise distance (PWD) countermeasure proposed in our previous work [21].

4.1. ASV systems and protocols

The ASV baseline systems used in this work, as well as the protocols and metrics, are identical to the ones defined in [21]. Accordingly we provide only a brief summary here.

Experiments were conducted with five different ASV systems: a standard GMM-UBM system with 1024 Gaussian components, a GMM-UBM system with factor analysis (FA) channel compensation according to [28] and three GMM supervector linear kernel (GSL) systems. They include a standard GSL system which applies a support vector machine classifier to GMM supervectors, a GSL system enhanced with channel compensation through nuisance attribute projection (GSL-NAP) [29] and a GSL system with FA supervectors (GSL-FA) [30]. They are all based on the LIA-SpkDet toolkit [31] and the ALIZE library [32] and are directly derived from the work in [30]. All systems use a common linear frequency cepstral coefficient (LFCC) parametrization extracted using SPro [33] with frames of 20ms duration and 10ms overlap.

All development was performed using the male subset of the 2005 NIST Speaker Recognition Evaluation dataset (NIST'05) whereas the male subset of the NIST'06 dataset is used for evaluation. The NIST'04 or NIST'08 datasets are used as background data, depending on whether the data is used for ASV or for spoofing purposes respectively. In all spoofing and countermeasure experiments, all impostor accesses are replaced with spoofed versions according to the algorithms described in Section 2.

4.2. Spoofing attacks and countermeasure setup

The setup for the voice conversion system is identical to [21], while for artificial signal generation we adopted the setup reported in [34]. Speech synthesis attacks were implemented using the voice cloning toolkit[1] with a default configuration. We used standard speaker-independent models provided with the toolkit which were trained on the EMIME corpus [35]. Synthesized speech is generated using the transcripts of the original impostor utterances.

While it is admittedly not representative of real scenarios, we assess countermeasure performance in a worst case scenario, where the attacker/spoofer has full prior knowledge of the ASV system. Voice conversion and artificial signal attacks thus use the same features used for ASV. We note that other work has observed only minor differences in vulnerability when the ASV systems used to effect spoofing are different [23]. Normalized features used in the LBP countermeasure are composed of 51 coefficients: 16 LFCCs and energy plus their corresponding delta and delta-delta coefficients.

The LBP countermeasure was implemented using the toolkit made publicly available by The University of Oulu[2]. Histograms of LBPs are created for all but the first and last rows of the textrogram, thereby obtaining a $58 \times (51 - 2) = 2842$ length feature vector.

4.3. Results

Results are illustrated in Table 2 for (a) voice conversion, (b) speech synthesis and (c) artificial signals. All values in Table 2 relate to the false acceptance rate (FAR) for a fixed false rejection rate of 10%. The baseline performance of the five ASV systems is illustrated in the second column of each table. The FA system gives the best performance with an FAR of 1%.

The effect of spoofing is assessed by replacing all impostor transactions with spoofing attacks. Results are illustrated in the third column of Table 2 (a)-(c). Significant degradations are observed in all cases, except for the artificial signal attacks and the three GSL-based systems. This is not a surprise since the GSL supervectors model speech at the GMM component level, whereas speech synthesis and artificial signal attacks target ASV systems at the feature level. The FAR for GMM-UBM and FA systems degrades significantly for all three attacks and voice conversion provokes consistent degradations in FAR for all five systems.

We now turn to countermeasure assessment which is reported first, independently and second, when combined with ASV. Figure 2 illustrates a detection error trade-off (DET) plot[3] for the LBP countermeasure and all three spoofing attacks. The equal error rate (EER) is 0% for artificial signal attacks, 0.5% for speech synthesis attacks and 8% for voice conversion at-

[1] http://homepages.inf.ed.ac.uk/jyamagis/software/page37/page37.html

[2] http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab

[3] TABULA RASA scoretoolkit: http://publications.idiap.ch/downloads/reports/2012/Anjos_Idiap-Com-02-2012.pdf

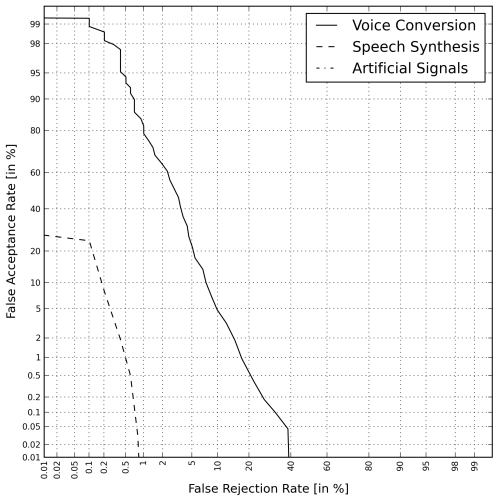

Figure 2: DET profiles illustrating LBP-countermeasure performance. The profile for artificial signals is not visible since the EER is 0%.

System	PWD	LBP
Voice Conversion	2.7	8
Speech Synthesis	10	0.5
Artificial Signals	35	0

Table 1: Comparison of performance in terms of EER (%) for the pair-wise distance (PWD) and new LBP-based countermeasures and the three different spoofing attacks.

tacks. A comparison to performance obtained with the PWD countermeasure reported in our previous work [21] is illustrated in Table 1 in terms of EER. While voice conversion still presents difficulties, the new countermeasure gives considerably better performance for both speech synthesis and artificial signals.

We now assess the impact of countermeasures on ASV performance. This requires the fixing of a countermeasure operating point. Here the threshold is once again fixed to a value which attains an FRR of 10%. The resulting FAR for all five ASV systems is illustrated in the fourth and fifth columns of Table 2 (a)-(c) for the PWD and LBP countermeasure respectively. Once again the threshold is set to give a fixed FRR of 10%. While the FAR for speech synthesis and artificial signal attacks is universally low, results confirm remaining vulnerabilities to voice conversion. This trend lies in contrast to results obtained for the PWD countermeasure (where performance is best for voice conversion) and thus fused countermeasure systems should be considered in future.

5. Conclusions and future work

This paper reports a new countermeasure for the protection of automatic speaker verification (ASV) systems from spoofing. The new countermeasure is based on the local binary pattern (LBP) analysis of sequences of acoustic vectors. Results show that the LBP countermeasure is less effective than previously reported solutions for voice conversion based spoofing attacks

System	Baseline	Spoof: ASV + —	PWD	LBP
GMM-UBM	6	77	2.3	6.2
GSL	6	88	2.6	7
GSL-NAP	3	84	2.5	6.7
FA	1	54	1.6	4.3
GSL-FA	2	82	2.5	6.6

(a) Voice Conversion.

System	Baseline	Spoof: ASV + —	PWD	LBP
GMM-UBM	6	82	8.2	0.8
GSL	6	35	3.5	0.3
GSL-NAP	3	27	2.7	0.3
FA	1	62	6.2	0.6
GSL-FA	2	20	2	0.2

(b) Speech Synthesis.

System	Baseline	Spoof: ASV + —	PWD	LBP
GMM-UBM	6	91	89	0
GSL	6	2	1.7	0
GSL-NAP	3	3	2.5	0
FA	1	75	72	0
GSL-FA	2	1	0.8	0

(c) Artificial Signals.

Table 2: ASV performance in terms of FAR (%) for the baseline (2^{nd} column) and under spoofing attacks without countermeasure (3^{rd} column) and with PWD and LBP countermeasures (4^{th} and 5^{th} columns respectively). All results relate to variable countermeasure thresholds which attain fixed FRRs of 10%.

but that better performance is achieved for previously unseen spoofing attacks which otherwise provoke significant increases in false acceptance. Being less reliant on prior knowledge, the work points to the potential for generalised countermeasures with greater practical value. Results also suggest that future work should consider fused approaches to countermeasure systems.

Even if generalised countermeasures have some potential, there is a clear need for formal spoofing and countermeasure evaluations. They should clearly differentiate the work of penetration testing / spoofing development from the far more important problem of countermeasure development. The latter should be conducted independently in a setting where the nature of spoofing attacks is unknown and varied. The development of effective countermeasures will then be extremely challenging.

6. Acknowledgements

This work was partially supported by the TABULA RASA project funded under the 7th Framework Programme of the European Union (EU) (grant agreement number 257289) and the ALIAS project (AAL-2009-2-049 - co-funded by the EC, the French ANR and the German BMBF).

7. References

[1] M. Blomberg, D. Elenius, and E. Zetterholm, "Speaker verification scores and acoustic analysis of a professional impersonator," in *Proc. FONETIK*, 2004.

[2] M. Farrús, M. Wagner, J. Anguita, and J. Hern, "How vulnerable are prosodic features to professional imitators?" in *Proc. Odyssey IEEE Workshop*, 2008.

[3] J. Lindberg and M. Blomberg, "Vulnerability in speaker verification - a study of technical impostor techniques," in *European Conference on Speech Communication and Technology*, 1999, pp. 1211–1214.

[4] J. Villalba and E. Lleida, "Speaker verification performance degradation against spoofing and tampering attacks," in *FALA workshop*, 2010, pp. 131–134.

[5] T. Masuko, T. Hitotsumatsu, K. Tokuda, and T. Kobayashi, "On the security of HMM-based speaker verification systems against imposture using synthetic speech," in *Proc. EUROSPEECH*, 1999.

[6] P. De Leon, M. Pucher, and J. Yamagishi, "Evaluation of the vulnerability of speaker verification to synthetic speech," in *Proc. Odyssey IEEE Workshop*, 2010.

[7] B. Pellom and J. Hansen, "An experimental study of speaker verification sensitivity to computer voice-altered imposters," in *Proc. ICASSP*, vol. 2, 1999, pp. 837–840.

[8] P. Perrot, G. Aversano, R. Blouet, M. Charbit, and G. Chollet, "Voice forgery using ALISP : Indexation in a Client Memory," in *Proc. ICASSP*, vol. 1, 2005, pp. 17 – 20.

[9] D. Matrouf, J. Bonastre, and J. Costa, "Effect of impostor speech transformation on automatic speaker recognition," *Biometrics on the Internet*, p. 37, 2005.

[10] T. Kinnunen, Z. Wu, K. A. Lee, F. Sedlak, E. S. Chng, and H. Li, "Vulnerability of Speaker Verification Systems Against Voice Conversion Spoofing Attacks: the case of Telephone Speech," in *Proc. ICASSP*, 2012, pp. 4401–4404.

[11] F. Alegre, R. Vipperla, N. Evans, and B. Fauve, "On the vulnerability of automatic speaker recognition to spoofing attacks with artificial signals," in *Proc. 12th EUSIPCO*, 2012.

[12] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *Biometrics, IET*, vol. 1, no. 1, pp. 3–10, 2012.

[13] M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori *et al.*, "Competition on counter measures to 2-d facial spoofing attacks," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–6.

[14] Z. Wu, E. Chng, and H. Li, "Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition," in *Proc. 13th Interspeech*, 2012.

[15] Z. Wu, T. Kinnunen, E. Chng, H. Li, and E. Ambikairajah, "A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. IEEE, 2012, pp. 1–5.

[16] P. De Leon, I. Hernaez, I. Saratxaga, M. Pucher, and J. Yamagishi, "Detection of synthetic speech for the problem of imposture," in *Proc. ICASSP*, 2011, pp. 4844–4847.

[17] A. Ogihara and A. Shiozaki, "Discrimination method of synthetic speech using pitch frequency against synthetic speech falsification," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 88, no. 1, pp. 280–286, 2005.

[18] P. De Leon, B. Stewart, and J. Yamagishi, "Synthetic speech discrimination using pitch pattern statistics derived from image analysis," in *Proc. 13th Interspeech*, 2012.

[19] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[20] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[21] F. Alegre, A. Amehraye, and N. Evans, "Spoofing countermeasures to protect automatic speaker verification from voice conversion," in *Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[22] J. Bonastre, D. Matrouf, and C. Fredouille, "Transfer function-based voice transformation for speaker recognition," in *Proc. Odyssey IEEE Workshop*, 2006, pp. 1–6.

[23] J.-F. Bonastre, D. Matrouf, and C. Fredouille, "Artificial impostor voice transformation effects on false acceptance rates," in *Proc. Interspeech*, 2007, pp. 2053–2056.

[24] J. Yamagishi, T. Nose, H. Zen, Z.-H. Ling, T. Toda, K. Tokuda, S. King, and S. Renals, "Robust speaker adaptive HMM based Text-to-Speech Synthesis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 17, no. 6, pp. 1208–1230, 2009.

[25] J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, and J. Isogai, "Analysis of Speaker Adaptation Algorithms for HMM-based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm," *IEEE transactions on Audio, Speech & Language Processing*, vol. 17, no. 1, pp. 66–83, 2009.

[26] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech communication*, vol. 27, no. 3, pp. 187–207, 1999.

[27] A. Roy, M. Magimai-Doss, and S. Marcel, "A fast parts-based approach to speaker verification using boosted slice classifiers," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 241–254, 2012.

[28] D. Matrouf, N. Scheffer, B. Fauve, and J.-F. Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification," in *Proc. Interspeech*, 2007.

[29] W. Campbell, D. Sturim, D. Reynolds, and A. Solomonoff, "Svm based speaker verification using a gmm supervector kernel and nap variability compensation," in *Proc. ICASSP*, vol. 1, may 2006, p. I.

[30] B. G. B. Fauve, D. Matrouf, N. Scheffer, J.-F. Bonastre, and J. S. D. Mason, "State-of-the-art performance in text-independent speaker verification through open-source software," *IEEE Transactions on Audio Speech and Language processing*, vol. 15, no. 7, pp. 1960–1968, 2007.

[31] J. Bonastre, N. Scheffer, D. Matrouf, C. Fredouille, A. Larcher, A. Preti, G. Pouchoulin, N. Evans, B. Fauve, and J. Mason, "Alize/spkdet: a state-of-the-art open source software for speaker recognition," in *Proc. Odyssey IEEE Workshop*, vol. 5, 2008, p. 1.

[32] J.-F. Bonastre, N. Scheffer, C. Fredouille, and D. Matrouf, "NIST'04 speaker recognition evaluation campaign: new LIA speaker detection platform based on ALIZE toolkit," in *NIST SRE'04*, 2004.

[33] G. Gravier, "Spro: speech signal processing toolkit," *Software available at http://gforge. inria. fr/projects/spro*, 2003.

[34] F. Alegre, R. Vipperla, and N. Evans, "Spoofing countermeasures for the protection of automatic speaker recognition from attacks with artificial signals," in *Proc. 13th Interspeech*, 2012.

[35] M. Wester, "The EMIME bilingual database," The University of Edinburgh, Tech. Rep., 2010.