Improving channel robustness in text-independent speaker verification using adaptive virtual cohort models

Andreas Nautsch*[1], Anne Schönwandt[2], Klaus Kasper[1], Herbert Reininger[2] and Martin Wagner[2]

[1]University of Applied Science Darmstadt, Department of Computer Science
[2]atip GmbH, Daimlerstraße 32, D-60314 Frankfurt am Main

Abstract: In speaker verification, score normalization methods are a common practice to gain better performance and robustness. One kind of score normalization is cohort normalization, which uses information about the score behaviour of known impostors. During enrolment, impostor verifications are simulated to get a speaker-specific set of the most competitive impostors (the cohort). In the present paper, one virtual cohort speaker is synthesized using the most competitive impostor's Hidden Markov Models (HMMs). These impostors are also users of the system and therefore their models have channel-specific information contrary to the universal background model, which provides channel- and speaker-independent models. On verification, cohort scores are obtained by an additional verification of the virtual cohort speaker. The cohort scores evaluate the candidate as an impostor. A cohort normalized score promises greater robustness.

This paper will study the effect of the introduced cohort normalization technique on the speaker verification system *atip VoxGuard*, which is based on mel-frequency cepstral coefficients and HMMs. *VoxGuard* can be used as either a text-dependent or a text-independent verification system. In this paper, emphasis is placed on text-independent speaker verification. Experiments using the *atip* speech corpus and the *SieTill* speech corpus showed improvements measured by the equal error rate on performance and robustness.

Index Terms — speaker verification; text-independent; cohort-based

1 Introduction

Reliable security approaches are becoming increasingly relevant. Especially for end-users of the commercial, financial, and government sectors, a robust, reliable, and secure verification is very important [Spe12]. Knowledge- or token-based solutions are thought to be very problematic. They can get lost or get passed on to a third party without authorization or even unintentionally [Sch05, Sie09]. In biometric systems, the user himself provides the basis for the security system. One topic of biometrics is speaker verification, which takes advantage of the uniqueness of the human voice in terms of the voicegram-obtained parameters of the speech signal such as the power spectrum [KL09, FC11].

*andreas.nautsch@stud.h-da.de

Speaker verification systems can be text-dependent (fixed pass phrase) or text-independent (unspecified pass phrase or free speech) [KL09]. Within text-independent speaker verification systems, it is possible to combine the advantages of both biometric and knowledge-based security systems: by using phoneme models as in (speaker-specific) speech recognition, a speaker can be scored depending on whether a randomized pass-phrase was correctly spoken. Otherwise, a replay attack could be assumed [Moh08].

A verification decision is based on a score rating the match for a user on a claimed identity. However, this score varies, due to intraspeaker variability (e.g., health conditions and speaking rates) and channel variability (acoustic speech signal) [BBF$^+$04, KL09, MT10, Sae11]. This paper focuses on channel variability. Score normalization methods are a common practice to gain a better performance and robustness.

Cohort-based score normalization methods use information about known impostors, such as the user-specific mean and the variance of impostor scores [SR05, HB05]. The approach introduced in this paper calculates a cohort score during speaker verification that provides channel specific information, unlike the universal background model (UBM). The virtual cohort models are synthesized from enrolled speaker models, which have channel-specific information (such as the telephone or microphone). Therefore this information is also placed in the cohort models.

2 System Overview

VoxGuard is a speaker verification system developed by atip GmbH. Delta mel-frequency cepstral coefficients (MFCCs) and acceleration MFCCs are used as features. These features are modelled by hidden Markov models (HMMs) representing the most common phonemes in the German language. Initially, a complete phoneme reference is provided as the UBM.

During enrolment, the speaker phoneme models are trained by adapting the distribution mean values of the UBM phoneme models, so that the features extracted from the enrolment sample are well-modelled. Re-enrolments are accomplished by an HMM-mixture of the former with new phoneme models.

During the verification process, two common comparison scores are computed: the score S_κ of the claimed identity κ and the score S_{UBM} of the UBM [Sch05, KL09, FC11, Sae11]. Both scores are similarity scores for estimating the comparison of the extracted feature vectors from an unknown person's utterance O and a reference ω modelling those vectors [KL09]:

$$S_\omega = P(O|\omega), \tag{1}$$

with ω consisting of phoneme models λ_ω.

As a text can be spoken in different sessions with different durations, a duration score normalization is calculated [Sch05] by dividing the logarithm of the score S_ω by the count of the feature vectors T (duration). According to the 2012 NIST speaker recognition

evaluation plan [NIS12] and common community approaches[1], the decision score is based on the log-likelihood ratio of the duration-normalized scores, see eq. 2 and 3.

$$S_{UBM}^{LLR} = \frac{1}{T} * \log \frac{S_\kappa}{S_{UBM}} \tag{2}$$

$$S_{UBM}^{LLR} = \frac{1}{T} * (\log P(O|\kappa) - \log P(O|UBM)) \tag{3}$$

3 Cohort Normalization

Cohort based score normalization methods can be applied to the scores of the most competitive cohort speakers for each target speaker's reference [Sch05]. A cohort is a group of impostors. In this paper, these impostors are considered to be enrolled speakers on the same application system. Hence, the impostor's HMMs are trained under the same channel influences as the target speaker's HMMs.

It is possible to specify a cohort model $\lambda_{\bar{\kappa}}$ for each target speaker's HMM λ_κ of a phoneme λ [Sch05]. A candidate can be scored by those models as a competitive impostor. Thus, if the candidate is scored higher by the cohort models than by the reference models, the candidate can be assumed to be a subversive user, or vice versa, the candidate can be assumed to be the user of the claimed identity himself, if the candidate is scored higher by the reference models than by the cohort models.

For the purpose of specifying cohort models, the n most competitive impostors are selected from a cohort corpus of size $m, n <= m$. To handle all cohort models, they can be synthesized into one virtual reference, implying the synthesis of a virtual cohort speaker.

3.1 Cohort speaker selection

According to Isobe and Takahashi [IT99], four selection methods for speaker verification have been introduced:

- speaker-based: the complete reference of a speaker is selected;
- phoneme-based: just the most competitive HMMs are selected;
- state-based: assumed HMMs are left-to-right HMMs, the most competitive HMM-states are selected (a selection due to a similar scored verbalization of a phoneme); and
- distribution-based: for each HMM, the most competitive distributions are selected.

[1] such as Kinnunen and Li [KL09], Campbell et al. [CCG+07], Poh and Kittler [PK08], Munteanu and Toma [MT10], and Isobe and Takahashi [IT99]

Phoneme-based cohort selection was chosen for this paper, because it had already been evaluated successfully in preliminary research. The phoneme-based cohort selection itself is performed by an impostor verification simulation. The most competitive impostors are then selected by the highest mean scores on each reference HMM.

In tab. 1, an exemplary extract from an impostor simulation is shown. For three speakers, the top three impostors by mean scores are presented on the phonemes of the German word *Wald* (phonetic script /v-a-l-t/).

Claimed identity	/v/		/a/		/l/		/t/	
	Spkr	$\mu_{S_{UBM}^{LLR}}$	Spkr	$\mu_{S_{UBM}^{LLR}}$	Spkr	$\mu_{S_{UBM}^{LLR}}$	Spkr	$\mu_{S_{UBM}^{LLR}}$
A	G	2.25	M	3.05	L	-0.67	N	0.34
	N	0.57	D	2.28	M	-0.81	K	0.15
	L	0.47	N	-0.98	D	-0.88	I	0.13
B	N	1.20	K	0.34	A	2.65	H	0.03
	M	-0.48	H	-1.78	K	-0.27	K	0.02
	M	-2.20	G	-2.08	N	-0.78	E	0.01
C	J	3.42	K	1.25	A	-0.45	E	0.62
	H	0.51	A	-1.14	E	-0.56	J	0.32
	L	-0.25	J	-2.41	M	-0.82	G	0.26

Table 1: Extract from phoneme-based impostor simulation

3.2 Synthesizing a virtual cohort speaker's reference

In cohort normalization, it is common to use multiple cohort speakers and cohort scores (e.g., [Sch05, HB05, KCD11, KKF06, Lon10, PMK09]). Isobe and Takahashi [IT99] introduced a virtual synthesized cohort speaker C_V, whose reference consists of the most competitive cohort models.

For each reference phoneme model λ_κ, n cohort phoneme models $\lambda_{\bar{\kappa}}^i$ can be selected. Thus the models of the most competitive impostor on phoneme λ can be considered as $\lambda_{\bar{\kappa}}^1$, the model of the second most competitve impostor as $\lambda_{\bar{\kappa}}^2$, ..., the model of the most uncompetitive impostor as $\lambda_{\bar{\kappa}}^m$.

In order to maintain one C_V model for each phoneme, a mixture of all the selected phoneme models is needed: in terms of n selected impostor models $\lambda_{\bar{\kappa}}^1, \lambda_{\bar{\kappa}}^2, \ldots, \lambda_{\bar{\kappa}}^n$ a cohort model $\lambda_{\bar{\kappa}}^V$ is synthesized by an HMM mixture of the selected models. Since this virtual cohort speaker contains all the information from the virtual speaker cohort, it is therefore taken as the *universal* virtual cohort speaker (Fig. 1). In this approach, only one additional verification on one cohort reference is done, because all competitive cohort references are mixed into one.

In fig. 1, the synthesis of a universal virtual cohort speaker C_V is given as an example for a selection quantity of $n = 2$. With speaker B as a reference, the selected cohort

speakers A, G, H, K, N and their mean scores are shown for each phoneme of the word *Wald*. The synthesis of the virtual cohort phoneme model $a_{\bar{B}}^{V}$ uses the selected phoneme models a_H, a_G of the cohort speakers H and G as the cohort phoneme models $a_{\bar{B}}^{1}, a_{\bar{B}}^{2}$ with regard to their rank of competitiveness. All virtual cohort phoneme models $\lambda_{\bar{B}}^{V}$ are stored in a databse and serve as a reference for the universal virtual cohort speaker.

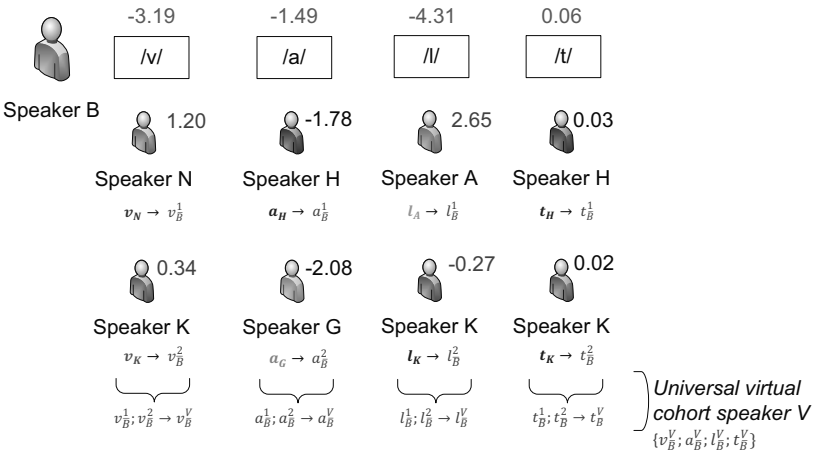

Figure 1: Synthesizing a universal virtual cohort speaker

Preliminary research showed that the selection of the two most competitive cohort phonemes performs best. When selecting a quantity of $n = 1$, there is overfitting to the most competitive virtual cohort speaker. This could be adjusted by a mixture of the phoneme models with the second most competitive cohort phoneme models. With $n > 2$, the cohort phoneme models also adapt to the uncompetitive cohort phoneme models and therefore the information value of the intended cohort score decreases.

In contrast to other common cohort-based approaches[2] using an impostor set of n cohort speakers, the approach introduced here relies on one universal virtual cohort speaker using $n = 2$ of the most competitive cohort speaker. Also, fewer resources are required during the verification as fewer score calculations need to be processed.

3.3 Score Normalization

A verification on the cohort reference produces the cohort score $S_{\bar{\kappa}}$. According to Türck [Tür08], the reference score S_κ is normalized by the cohort score using the log-likelihood ratio:

$$S_{\bar{\kappa}}^{LLR} = \frac{1}{T} * \log \frac{S_\kappa}{S_{\bar{\kappa}}}, \tag{4}$$

[2] e.g., in Hébert and Boies for text-dependent speaker verification [HB05], Sturim and Reynolds [SR05], Poh and Kittler [PK08], Poh et al. [PMK09], and Karam et al. [KCD11]

where, differing from eq. 2, the cohort score is used instead of the UBM score. To gain the advantages of both LLRs, a score fusion is carried out by adding both LLR scores using equal weighting:

$$S_{UBM+Cohort} = S^{LLR}_{UBM} + S^{LLR}_{\bar{\kappa}}. \tag{5}$$

4 Experimental Results

The approach described in section 3 was evaluated on the *atip* speech corpus with 27 speakers (18 male, 9 female) and on a part of the *SieTill* speech corpus with 356 speakers[3] (188 male, 168 female). All speakers of the *atip* speech corpus were recorded in a quiet room with the *AT4033* microphone in one session. Speakers of the *SieTill* speech corpus were recorded by telephone (ISDN).

The *atip* speech corpus was introduced by Kunz et al. [KKR+11] for evaluating continuous text-independent speaker verification, and it was extended during preliminary research. Therefore, phonetically balanced sentences were important, hence the *Nordwind und Sonne* fable (see [Pet99]) and the story *Buttergeschichte* (see [Pet99]) were used, because they are standard phonetic and linguistic texts respecting phonetic balance within the texts. As a third text, a part of an online newspaper article[4] was read[5], because it is assumed to be more like free speech. This excerpt is referred to as the *Mainufer* text in this paper. Each text was freely spoken. In this evaluation setup, each speaker of the *atip* speech corpus was enrolled with the *Nordwind und Sonne* fable. The impostor simulation was performed using the story *Buttergeschichte*, and for verification, the *Mainufer* text was used. Each enrolled speaker of the *atip* speech corpus was used for cohort selection and evaluation[6].

The *SieTill* speech corpus contains three numbers in the range of $0, 1, \ldots, 9$. This corpus was divided into a part with 56 speakers (35 male, 21 female) used for cohort selection and 300 speakers used for evaluation[7].

The approach introduced was evaluated using the equal error rate (EER) metric in terms of the false match rate (FMR) and the false non-match rate (FNMR). The performances of LLR-UBM, LLR-Cohort, and UBM+Cohort are compared in fig. 2 by detection–error tradoff (DET) graphs, and in tab. 2, by EERs. In both speech corpora, the fusioned UBM+Cohort outperforms LLR-UBM and LLR-Cohort in terms of EER.

VoxGuard reaches an EER of 6.64% on the *SieTill* speech corpus with the LLR-UBM ap-

[3] The corpus was separated by speaker id, the last 356 speakers were used for testing.

[4] see: `http://www.focus.de/panorama/welt/dauerregen-frankfurt-ruestet-sich-gegen-das-hochwasser_aid_589684.html` (Focus Online, 26.05.2012)

[5] segmented into 26 speech samples, with durations from one to eleven seconds

[6] The setup used for the *atip* speech corpus did not conform to a strict interpretion of ISO/IEC 19795-1 clause 7.6.3.2.1 [ISO05], since cohort models are based upon impostor models. This was necessary, due to the limited data of this speech corpus. The application scenario of this setup is a simulation of known users subversively claiming another user's identity as their own, in order to defraud the other user.

[7] About 10 samples per speaker, every sample had a duration of 2.784 seconds

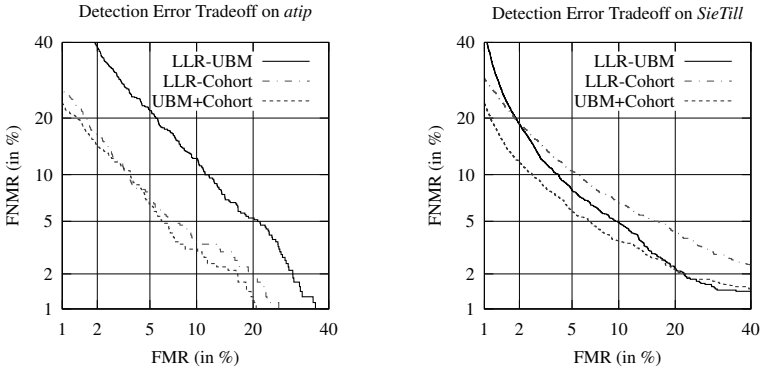

Figure 2: DET graphs comparing LLR-UBM, LLR-Cohort, and LR-UBM-Cohort, left: on the *atip* speech corpus, right: on the *SieTill* speech corpus

speech corpus	*atip*	*SieTill*
LLR-UBM	10.86	6.64
LLR-Cohort	5.98	7.76
UBM+Cohort	**5.56**	**5.48**

Table 2: Speaker verification performances comparison by EER (in %)

proach. On the *atip* speech corpus, there was a higher EER of 10.86%. The aim was to create an approach that performs approximately equally. This approach should be independent of the speech corpus used and of the recording channels. Because a channel adaption of the UBM is not easy with a small amount of speech data, the cohort normalization approach introduced was used to achieve more robustness against the channel variability and have an approximately equal performance on both speech corpora.

By using the LLR-Cohort approach, the EER decreases significantly on the *atip* speech corpus, to 5.98%. Though more robustness was found on the *atip* speech corpus by using only the LLR-Cohort approach, the evaluation on the *SieTill* speech corpus could not confirm this. The EER increased by 1.12 percentage points. At the EER-threshold, 3.90% of the samples were better classified by the LLR-Cohort score than by the LLR-UBM score, but on the other hand, there were 4.95% of the samples that were worse classified, and 2.68% of the samples were classified incorrectly by both scoring approaches. By fusing the LLR-UBM and the LLR-Cohort, the UBM+Cohort approach has the advantages of both and outperforms them, with an EER of 5.48%. On the *atip* speech corpus, the best performance was also observed with the UBM+Cohort approach, with an EER of 5.56%.

Altogether, the fusion of LLR-UBM and LLR-Cohort, UBM+Cohort, outperforms both approaches on both evaluation copora. The UBM+Cohort approach reached approximately the same EER on both speech corpora. If the classification using the LLR-UBM approach is not optimal, improvements could be gained by using the score fusion introduced with the LLR-Cohort score. Hence, the UBM+Cohort approach improves the robustness of *atip VoxGuard*.

To examine the score behaviours of S_{UBM}, S_{κ} and $S_{\bar{\kappa}}$ the samples were analysed regarding continuous scores. In the following, samples from the *SieTill* speech corpus are analysed that were classified better or worse by LLR-Cohort than by LLR-UBM. It turned out that the verification phrase *eins null drei* $(1, 0, 3$, phonetic script /ʔ[aɪ]ns nʊl dʁ[aɪ]/), accounts for 18.12% of the worse classifications by LLR-Cohort.

In fig. 3, a correct verification by LLR-Cohort is presented in order to point out the advantages of this approach. The scores of a verification trial are shown, where the logarithm scores of the UBM, a target speaker's models, and the target's cohort models are assigned to the continuous observation of an impostor's utterance. In this example, the LLR-UBM accepts the impostor as genuine, with $S_{UBM}^{LLR} = 0.71$. The subversive trial is rejected by the LLR-Cohort approach, with $S_{\bar{\kappa}}^{LLR} = -1.87$. Overall, the UBM+Cohort rejected the trial with $S_{UBM+Cohort} = -1.16$ as well. This indicates a user-adaptive perfomance increase and a higher channel robustness by outperforming the UBM, which is assumed to be channel-independent, having no channel-specifc information.

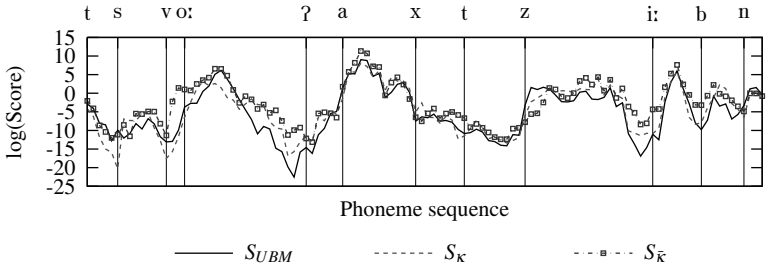

Figure 3: Logarithm scores of S_{UBM}, S_{κ} and $S_{\bar{\kappa}}$ on continuous observation of an impostor utterance with the phrase *zwo acht sieben* (2, 8, 7, phonetic script /tsvo: ʔaxt zi:bn/)

Analysing the incorrectly classified samples from the LLR-Cohort approach, it can be observed that some cohort models have extreme values, while the UBM and reference scores do not. Fig. 4 shows another impostor trial with the phrase *eins null drei*. Overall, the UBM scores outperform the cohort scores, because the cohort scores are very low for specific phonemes. In this example, the cohort scores have outliers on phonemes /l/, /d/, and /ʁ/, and furthermore, a better performance of the UBM on the phonemes /ʔ/, /s/, and /[aɪ]/.

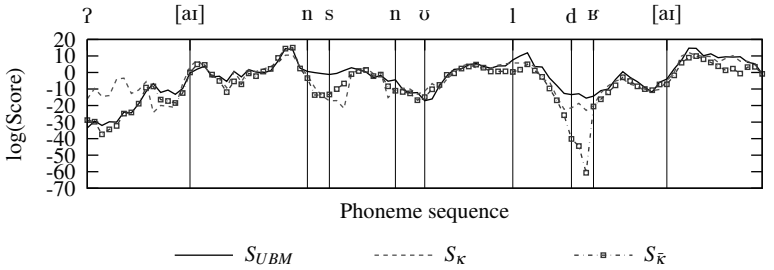

Figure 4: Outlying cohort scores on continuous impostor utterance observation

5 Conclusion and Future Research

This paper introduced a cohort normalization approach for improving the channel robustness of the HMM-based speaker verification system *atip VoxGuard*. For every user, a virtual cohort model was synthesized by mixing the corresponding phoneme models of the two most competitive impostors from a cohort corpus. This cohort corpus consisted of users who were enrolled subject to the same influences of channel and noise. All virtual cohort models together represented a universal virtual cohort speaker that modelled the most competitive (in the context of one specific user) impostors. During a verification, one cohort score was calculated by using the cohort phoneme models of this universal virtual cohort speaker.

It was shown that this cohort score cannot replace the UBM for verification for all channels and achieve more robustness by channel-adaption. Since the UBM and cohort approaches do not perform consistently on different speech corpora with different input channels, a fusion that combined both approaches outperforms them and is more likely to be stable.

The presence of cohort score outliers points to a problem; further researches on phoneme model training or cohort model synthesis might solve this problem. In order to obtain more discriminative target speaker models and cohort models, the training of the UBM might be examined, because both depend on it, whether directly or indirectly. According to Hasan and Hansen [HH11], the selective use of speaker data for UBM construction promises a higher performance. With the intention of increasing the security of mobile phones, a possible future application could make use of continuous speaker verification: 'Additional protection against intruders can be given if voice verification is made concurrent to phone calls' [KKR+11]. It still needs to be clarified whether the proposed cohort normalization approach can produce a performance and robustness gain in continuous speaker verification. In the future, a test of the approach introduced in the present paper could be carried out by taking part in the NIST speaker recognition evaluation.

References

[BBF+04] Frédéric Bimbot, Jean-François Bonastre, Corinne Fredouille, Guillaume Gravier, Ivan Magrin-Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega-García, Dijana Petrovska-Delacrétaz, and Douglas A. Reynolds. A Tutorial on Text-Independent Speaker Verification. In *EURASIP Journal on Applied Signal Processing*, pages 430–451, 2004.

[CCG+07] W.M. Campbell, J.P. Campbell, T.P. Gleason, D. A. Reynolds, and Wade Shen. Speaker Verification Using Support Vector Machines and High-Level Features. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 2085–2094, 2007.

[FC11] A. Fazel and S. Chakrabartty. An Overview of Statistical Pattern Recognition Techniques for Speaker Verification, 2011.

[HB05] M. Hébert and D. Boies. T-Norm for Text-Dependent Commercial Speaker Verification Applications: Effect of Lexical Mismatch. In *Acoustics, Speech, and Signal Processing (ICASSP)*, pages 729 – 732, 2005.

[HH11] Taufiq Hasan and John H. L. Hansen. A study on Universal Background Model training in Speaker Verification. *Audio Speech and Language Processing, IEEE Transactions on*, pages 1890–1899, Sep. 2011.

[ISO05] ISO. Text of ISO/IEC FDIS 19795-1, Information technology - Biometric performance testing and reporting - Part 1: Principles and framework, 2005. ISO-19795-1.

[ISO12] ISO and IEC. Information technology – Vocabulary – Part 37: Harmonized biometric vocabulary, 2012. Draft ISO/IEC DIS 2382-37.

[IT99] T. Isobe and J. Takahashi. A new cohort normalization using local acoustic information for speaker verification. In *Acoustics, Speech, and Signal Processing (ICASSP)*, pages 841–844, 1999.

[KCD11] Z.N. Karam, W.M. Campbell, and N. Dehak. Towards reduced false-alarms using cohorts. In *Acoustics, Speech and Signal Processing (ICASSP)*, pages 4512 – 4515, 2011.

[KKF06] T. Kinnunen, E. Karpov, and P. Franti. Real-Time Speaker Identification and Verification. In *Audio, Speech, and Language Processing*, pages 277 – 288, 2006.

[KKR+11] Max Kunz, Klaus Kasper, Herbert Reininger, Manuel Möbius, and Jonathan Ohms. Continuous Speaker Verification in Realtime. In *BIOSIG 2011 - Proceedings of the Special Interest Group on Biometrics and Electronic Signatures*, pages 79–88, 2011.

[KL09] Tomi Kinnunen and Haizhou Li. An overview of text-independent speaker recognition: from features to supervectors, 2009.

[Lon10] Chris Longworth. *Kernel Methods for Text-Independent Speaker Verification*. PhD thesis, University of Cambridge, 2010.

[Moh08] Aanchan K. Mohan. Combining speech recognition and speaker verification. Master's thesis, New Brunswick Rutgers, The State University of New Jersey, 2008. URL: `http://hdl.rutgers.edu/1782.2/rucore10001600001.ETD.17528`.

[MT10] D.-P. Munteanu and S.-A. Toma. Automatic Speaker Verification Experiments using HMM. In *2010 8th International Conference on Communications (COMM)*, pages 107–110, 2010.

[Nau12] Andreas Nautsch. Kohortenbasierte Score-Normierung zur robusteren textunabhängigen Sprecherverifikation. B.Sc. thesis, University of Applied Science Darmstadt, 2012.

[NIS12] NIST – National Institute of Standards and Technology. The NIST Year 2012 Speaker Recognition Evaluation Plan. Web: `http://www.nist.gov/itl/iad/mig/upload/NIST_SRE12_evalplan-v11-r0.pdf`, 2012.

[Pet99] Benno Peters. Prototypische Intonationsmuster in deutscher Lese- und Spontansprache. In *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*, pages 1–171, 1999.

[PK08] N. Poh and J. Kittler. Incorporating Model-Specific Score Distribution in Speaker Verification Systems. In *IEEE Transactions on Audio, Speech, and Language Processing*, pages 594–606, 2008.

[PMK09] Norman Poh, Amin Merati, and Josef Kittler. Adaptive Client-Impostor Centric Score Normalization: A Case Study in Fingerprint Verification. In *International conference on Biometrics: Theory, applications and systems*, pages 245–251, 2009.

[Sae11] Rahim Saeidi. *Advances in Front-end and Back-end for Speaker Recognition*. PhD thesis, University of Eastern Finland, 2011.

[Sch05] Holger Schalk. *Biometrische Authentifikation auf Basis von Sprache unter Verwendung stochastischer und signalorientierter Modelle*. PhD thesis, Johann Wolfgang Goethe Universität Frankfurt am Main, 2005.

[Sie09] Ingo Siegert. Implementierung einer Sprecherverifikation für ein generisches Telefon-Dialogsystem. Master's thesis, Otto-von-Guericke-Universität Magdeburg, 2009.

[Spe12] Speech Technology Magazine. Voice Biometrics Poised for Growth, Voice authentication and transaction verification provide robust security. SpeechTechMag.com: `http://www.speechtechmag.com/Articles/News/Industry-News/Voice-Biometrics-Poised-for-Growth-80071.aspx`, 2012.

[SR05] D. E. Sturim and D. A. Reynolds. Speaker adaptive cohort selection for tnorm in text-independent speaker verification. In *in Proc. ICASSP*, pages 741–744, 2005.

[Tür08] Ulrich Türck. *Compensation Techniques for Network Mismatch in Telephone-Based Speaker Verification*. PhD thesis, Technischen Universität München, 2008.