Age-Related Voice Disguise and its Impact on Speaker Verification Accuracy

Rosa González Hautamäki, Md Sahidullah
Tomi Kinnunen, Ville Hautamäki

School of Computing
University of Eastern Finland, Joensuu, Finland
{rgonza, sahid, tkinnu, villeh}@cs.uef.fi

Abstract

This study focuses in the impact of age-related intentional voice modification, or age disguise, on the performance of automatic speaker verification (ASV) systems. The data collected for this study includes 60 native Finnish speakers (29 males, 31 females) with age range between 18 and 73 years. The corpus consist of two sessions of read speech per speaker. Our experiments demonstrate vulnerability of modern ASV systems when a person attempts to conceal his or her identity, by modifying the voice to sound like an old or young person. For our i-vector PLDA system, the increase in equal error rate (EER), in the case of male speakers, was 7-fold for the attempt of old voice and 11-fold for young voice. Similar degradation in performance is observed for female speakers with a 5-fold increase in EER for old voice disguise and a 6-fold increase for young voice disguise. We further analyze the factors affecting the performance of ASV systems for the studied speech data. In our experiments, male speakers were found more successful in disguising their voices. The effect on fundamental frequency ($F0$) was also studied. The mean $F0$ distributions showed a shift towards higher frequencies when speakers attempted a young voice, which relates to the perception that younger speakers' $F0$ values tend to be higher than for older speakers.

1. Introduction

Biometric authentication systems measure physiological or behavioral characteristics of a person to verify his or her identity [1]. Known biometric traits include fingerprint, DNA, palmprint, hand geometry, earlobe geometry, iris, retina, voice and signature patterns. They are considered unique personal identifiers and some are used in a number of authentication applications. Nowadays, many personal identity documents, such as passports, include biometric identifiers.

Among the factors that affect the performance of biometric systems are variations in the acquisition method, environment and changes in traits themselves caused by physiological or pathological conditions [1, 2]. *Permanence* or sufficient invariance of a trait over time is desirable. According to biometric characteristic comparison, DNA, iris, ear shape, fingerprints and palmprints show high level of permanence over time, while gait, keystroke and voice characteristics vary over time [3]. *Ageing* is a factor that affects the permanence of a biometric over time.

The ageing process causes changes in physical and behavioral characteristics of a person. The effects of ageing can be direct or indirect [3]. The effects of *direct* ageing relate to the person's age and the natural process of growth. In contrast, *indirect* ageing refers to changes caused by injuries or health related conditions considered to develop with age. Effects of ageing include the loss of elasticity, growth, wearing, muscle atrophy and reduced ability. They affect the appearance of physical characteristics, preventing or disturbing the acquisition and the generation of features that represent the trait [3, 2]. Direct ageing effects could be observed in the skin elasticity, for example, for face, fingerprints and palmprints. Other effects of direct ageing include body movements affected by reduced strength in muscle, reduced hearing and sight. The human voice experiences a reduced lung capacity, vocal muscle atrophy and degeneration [4, 5].

Ageing-related changes in the human voice are obvious during adolescence and the elderly, but ageing effects are present in the voice through adulthood too. The main cause of change is physiological, the degeneration and growth of the speech production apparatus, general health and lifestyle contribute to vocal change also, hence many studies consider vocal ageing as an individual process. Ageing affects the speech organs in the same way as many other bodys organs: weakening of muscles and loss of tissue elasticity [4, 5].

It is known that, besides the well-studied effects of communication channel and environment, the performance of automatic speaker verification systems is greatly affected by variations in the speaker's voice characteristics. These variations in the speaker's voice can be classified as *unintentional* or *intentional* [6]. Unintentional variations are caused by the emotional or physical state of the speaker, such as sore throat, biological ageing, etc. Concerning the intentional variations, two cases can be identified: firstly, the speaker may attempt to be identified as another person (impersonation). Secondly, the speaker may attempt to conceal his or her identity to prevent from being identified. We will refer to this latter intentional modification as voice *disguise*. In the field of forensics, a suspect could attempt disguising his or her identity by voice modification, or to conceal identifiable characteristics like age, gender and dialect.

A few studies have investigated the impact of voice ageing to speaker recognition accuracy [7], as well as considering the speaker's age as a factor affecting system performance evaluation [8, 9]. In [7], the impact of long-term ageing on the performance of two automatic speaker verification systems (ASV) was evaluated. It was reported that age difference between enrollment and test samples increased equal error rate (EER) in the case of male speakers from 4.61%, when the difference was less than a year, to 32.74% when the age difference range was 51-60 years. The authors suggest that an age compensation approach should be included in ASV to deal with age-related variations.

In [8], the age factor is considered in the performance of an automatic system by analyzing the age difference between target and non-target speakers. Age difference ranges were defined so that the number of target speakers was approximately the same for all age ranges. The effect was measured in terms of gender-dependent false alarm probability at a given miss probability of 1% when the age difference between the target and non-target increases. The false alarm probability was reported to reduce when the age difference is 5-9 years compared to difference of 0-4 years. It was further reduced for age differences exceeding 40 years. The goal of the study was to introduce the concept of population demography effect in the evaluation of speaker recognition systems.

In [9], the role of the speaker's age in the acoustic space was studied. The role of age in the acoustic space was developed from the known age information in the data. The study indicated that the performance of speaker identification systems can be improved by developing age-sensitive eigenvoice models in the acoustic space. These studies suggest that speaker recognition systems should take the age factor into account to increase recognition accuracy.

Most of the age-related studies cited above assume a *natural* biological ageing process that takes place when the time lapse between enrollment and test phase increases. In contrast, in this study, we address the problem of *intentional* age modification in the speaker's voice. Such situation could take place, for instance, in public places when a person pretends a different age when facing social exposure, for example in bars and restaurants. Further, social network applications that make use of videos and voice recordings are open to situations in which a user may target a fake age. In addition, call center applications without face-to-face contact are prone to intentional modification to conceal one's age, specifically in the case when demographic restriction or discrimination is expected, for example minors accessing gambling services or elderly requesting information services.

We study the performance of six automatic speaker verification systems based on GMM-UBM and on i-vectors, with focus on age-related voice disguise. We describe the data collection protocol and the experimental design. Finally, we present an analysis of factors affecting the accuracy of the systems using the self-reported speaker information, such as, age, gender and experience in voice disguise. We also include an analysis of the *fundamental frequency* $F0$ feature which often plays a key role in voice modification attempts.

2. Data and evaluation protocol

The data collected for this study consists of age disguise as the intentional modification of the speakers' voice without using any auxiliary objects such as blocking one's mouth, nostrils or the recording device. The main instruction given to the participants was *to not sound like themselves*. The speech data was collected under controlled conditions in the same silent office environment for all the speakers. The speakers are native Finnish speakers, and the corpus consists of reading sentences and spontaneous speech from free selected topics.

2.1. Recording conditions

Two sessions are recorded for each speaker in two different days. The audio was collected using Zoom H6 Handy recorder, a portable device that allows capturing high quality audio using multiple microphones, at a sampling rate of 44.1 kHz and 32 bits precision. For this corpus, an omnidirectional headset microphone (Glottal Enterprises M80) was connected to the recorder, along with an Electroglottograph (EG2-PCX2) to record glottal activity besides the acoustic microphone data. In addition, parallel recording was carried out by voice recording apps in two smartphones: Nokia Lumia 635 and Samsung Galaxy Trend 2. In this study, we make use of audio data from the head mounted close-talk microphone and the two smartphones only.

Each participant performed 10 tasks per session, including: spontaneous speech, reading text and impersonation tasks. For the purpose of this study, we consider material collected from the **reading text** for 3 tasks only. It consists of reading two phonetically balanced texts, with a total of 11 sentences in Finnish language and 2 sentences in English, as illustrated in Fig. 1. One task consisted of reading in the speaker's modal voice without any intentional modifications, and two tasks consisted of modifying the voice to sound like an old person and a young person (e.g. a child).

The text material included the Finnish version of the "The Rainbow passage" and "The north wind and the sun" (See Appendix A), plus two TIMIT sentences, SA1 and SA2, in English: "She had your dark suit in greasy wash water all year" and "Don't ask me to carry an oily rag like that".

2.2. Summary of the participants

A total of 60 speakers were recorded, with almost perfect gender balance with 31 females and 29 males. The age range is from 18 to 73 years. The speakers provided also the following additional information: English proficiency, other languages spoken, profession, educational level, place of birth, residence place during elementary education, dialect, experience in voice modification, smoking habits and other free-worded information that could affect the voice quality and performance of the tasks. All the participants were adults (18+ years old) and signed a consent form to allow the use of their speech data for research purposes.

2.3. Audio segmentation and post processing

Each session was recorded into a long audio file without interruptions. Manual segmentation was carried out to produce 39 segments per session (13 sentences in 3 voice modes). Four simultaneous recording were segmented by annotating the time stamp in seconds for the beginning and end of each task and sentence in seconds. This annotation was then used to automatically cut the long recordings into sentence long segments. The data was downsampled to 8kHz to match the sampling rate of the development data. Figure 1 shows a description of the speech data per speaker in a single session.

3. Recognition experiments

We consider a text-dependent recognition setup where the target and impostor speakers have lexically matched contents. The reading material for natural voice in session 1 is used for enrollment, resulting in 13 phrase-dependent adapted models per speaker. In the testing phase, genuine trials correspond to the target speaker's second session and, for impostors, the reading text from both sessions to increase the number of non-target speaker trials. We considered the *natural* voice test as the baseline and *disguise* case for the segments corresponding to old and young voices. We explore the case of channel mismatched by substituting the test segments with the segments recorded with the two smartphones, while enrollment segments in all the cases correspond to audio from the close-talk microphone. The summary of the trials is shown in Table 1.

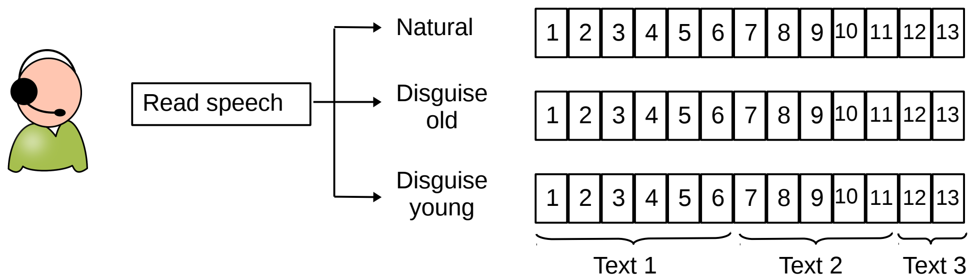


Figure 1: Diagram of the speech audio for one session used for this study. The blocks represent segments for each sentence in the text, the Finnish version of "The Rainbow passage" (Text 1), "The north wind and the sun" (Text 2) and two TIMIT sentences (in English Text 3).

Table 2: Description of development data and setup for the automatic speaker verification systems in this study.

		UBM training data (512 Gaussians)	**T-matrix**	**Scoring**
GMM-UBM	System1	NIST SRE 04, 05, 06, Switchboard and FISHER	—	LLR
	System2	NIST SRE 04, 05, 06, Switchboard, FISHER, PERSO	—	LLR
i-vector	System3	NIST SRE 04, 05, 06, Switchboard and FISHER	Same as UBM	PLDA
	System4	NIST SRE 04, 05, 06, Switchboard and FISHER, PERSO	Same as UBM	PLDA
	System5	NIST SRE 04, 05, 06, Switchboard, FISHER	Same as UBM	Cosine
	System6	NIST SRE 04, 05, 06, Switchboard, FISHER, PERSO	Same as UBM	Cosine

Table 1: Description of test trials for text-dependent speaker verification used in this study.

		Trials	
	Test	**Genuine**	**Impostor**
Male	Natural voice	377	21112
	Disguise old	377	21112
	Disguise young	377	21112
Female	Natural voice	403	24180
	Disguise old	403	24180
	Disguise young	403	24180

3.1. Speaker recognition systems

We evaluate the age disguise effect using two speaker verification systems based on Mel-frequency cepstral coefficient (MFCC) features: a *Gaussian mixture model with universal background model* (GMM-UBM) [10] and an *i-vector* based system. The scoring of the extracted i-vectors is performed using both cosine scoring [11] and probabilistic linear discriminant analysis (PLDA) [12]. We used i-vector dimensionality of 400 and 200 speaker factors with simplified PLDA.

Table 2 describes the developmental setup and data used to train each system. Besides including data from NIST Speaker Recognition Evaluation (SRE) corpora, Switchboard and FISHER, a corpus in Finnish language called *PERSO* [13] is used. PERSO contains audio material from 66 speakers, including 33 male and 33 female. It contains approximately 43,000 utterances, 40 minutes of continuous speech in *small* part and approximately 4 hours in *long* part. The corpus was collected to create more appropriate speech synthesis options for text-to-speech (TTS) conversion applications as seen in assisted communication devices. Our goal was to include a Finnish language corpus to add more variation to the ASV system's component considering all our test data is in Finnish language.

4. Results

Table 3 presents the performance of our ASV systems in terms of equal error rate (EER) which corresponds to equal miss and false alarm rates. We calculate the EER using the implementation for ROC convex hull method (ROCCH-EER) in the BOSARIS Toolkit [14]. Comparing the baseline (natural voice) to age disguise (old and young voices), we observe a dramatic increase in EER in the presence of age-related disguise, especially when the voice is modified to sound like a younger person. Previous studies such as [5] support that young voices imply a higher fundamental frequency or perceived pitch, and it is considered to be a challenging modification for ASV systems that use modal voice for enrollment.

In the case of channel mismatch, where the enrollment corresponds to the natural voice from the close-talk microphone and the test sample from audio recorded with one of the two smartphones, the results are consistent with the same channel test. We observe an increase on EER for mismatched channel data for old voice and a higher increase for young voice in all the systems. Comparing the performance of both smartphones, we observed in most of the cases that Samsung Galaxy data had slightly lower EER than the Nokia Lumia smartphone. In general, the differences between the smartphones are small, and would be caused by differences in casing, microphone quality and recording application.

Including the Finnish corpus (PERSO) in the UBM training reduced the EER for GMM-UBM system (System2). For the i-

Table 3: Performance in terms of equal error rate (EER,%) for GMM-UBM (System1-2) and i-vector systems (System3-6) for female and male speakers with natural voice and two disguised voices: Old and Young. Channel variation results are also included.

		Microphone			**Samsung Galaxy Trend 2**			**Nokia Lumia 635**		
		Natural	**Disguise old**	**Disguise young**	**Natural**	**Disguise old**	**Disguise young**	**Natural**	**Disguise old**	**Disguise young**
Female	System1	10.13	28.45	37.63	9.40	27.03	36.29	9.65	27.31	36.86
	System2	6.88	25.41	35.45	7.52	26.37	35.84	8.86	26.50	36.31
	System3	5.05	24.38	31.68	8.92	27.95	32.85	9.44	29.61	34.22
	System4	7.13	27.71	34.98	11.48	29.13	35.49	12.13	28.72	35.55
	System5	6.92	25.63	33.90	10.79	26.92	36.30	10.88	28.58	36.26
	System6	10.38	29.28	37.65	13.10	29.80	39.51	14.18	31.75	38.99
Male	System1	4.48	21.66	31.40	4.99	23.89	30.60	5.93	25.04	31.31
	System2	4.08	20.55	30.57	5.27	22.84	30.26	5.48	24.30	31.56
	System3	2.82	19.45	30.10	4.29	20.16	30.50	5.47	21.08	32.08
	System4	3.27	19.84	31.66	6.02	22.00	31.30	5.13	23.00	32.42
	System5	2.71	20.79	31.19	3.88	21.78	30.88	4.98	23.28	32.48
	System6	5.14	23.83	35.00	7.96	24.69	33.66	7.13	25.09	35.20

vector systems, including the Finnish data in UBM and T-matrix training did not improve the performance, as it can be seen in the results for System4 and System6. This may be caused by the relatively small number of speakers in the corpus compared to number of English speakers in the rest of the development set.

Figure 2 further shows the detection error trade-off (DET) curves for natural voice and the two disguised voices (old and young).

We calibrated the scores for System3 using logistic regression with target probability 0.5 and false alarm and miss set to 1, following [15], then the decision threshold is set at origin. Table 4 shows a degradation for the false rejection rates (FR %) when voice disguise is introduced for matched channel case, using the close-talk microphone data only.

Table 4: *False rejection rates (FR %) for i-vector PLDA (System3) after calibration of scores with decision threshold set at origin for close-talk microphone data.*

Test	**Female**	**Male**
Natural	5.21	2.65
Disguise old	26.05	20.69
Disguise young	34.74	33.16

5. Age disguise affecting factors for speaker verification

To analyze the effect of age-related disguise in our ASV systems accuracy, first, we study the changes in fundamental frequency ($F0$) when age disguise is present, and second, we evaluate the the speakers' self-reported information to identify the factors that have an effect on successful age disguise.

The $F0$ was extracted using a function from VoiceSauce [16] that uses Praat [17] to estimate this feature with the autocorrelation method [18]. We then compute the mean $F0$ for every utterance. Figure 3 shows the distribution of mean $F0$ for the natural voice along the two disguise cases separately for both genders. We observe an increase on the mean $F0$ values for young disguise, especially for female speakers. This may reflect that females tend to vary their $F0$ with relative ease, and by increasing it fit to the perception of a younger speaker voice.

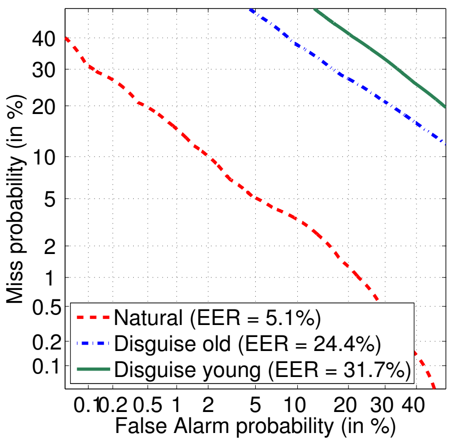
We also studied the correlation between the self-reported speaker information and age disguise effect on the ASV systems accuracy. The goal was to find if there are any factors that could explain *why* some speakers were able to spoof the ASV system better than others. To accomplish this goal, first, scores for System3 were turned into hard decisions and second, the decisions of genuine (same-speaker) trials were used to measure the effect of the self-reported speaker information. To turn System3 scores to decisions, we used logistic regression to find the scale and shift of the scores as explained in [15]. We fit the logistic regression model with prior probability of observing a genuine trial equaling to 0.5 and false alarm and miss costs set to 1 as in Table 4. We optimized the scale and shift in the same set of scores so that the results can be seen as the best possible (oracle) decisions.

To continue the analysis, our dependent variable is *success*, so that when a speaker succeeds in disguise his or her voice, *success* is set as *true*, otherwise as *false*. We define the speaker information factors *age, gender, English proficiency, education level, experience in voice disguise* and *smoking habits* as the predictors in our model. The goal is to predict success using the above mentioned factors by fitting a logistic regression model separately for old and young voice disguise experiments.

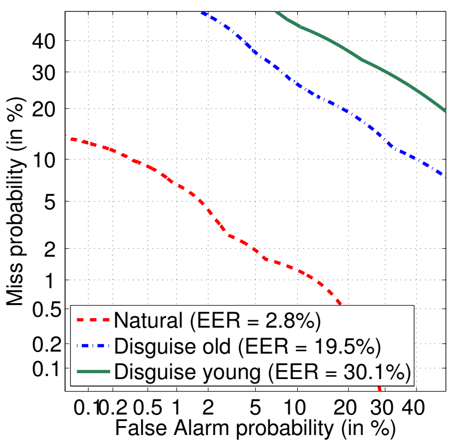
The results indicate that for our data age was not a significant factor in neither old or young voice disguise case, in other words, no age group was significantly better at disguise than others. In our experiments, English proficiency (beginner, intermediate and advanced) was not a significant factor either.

Interestingly, by analyzing the gender factor, we notice that males had a regression coefficient of 1.537 and p-value < 0.0001 for young voice disguise case and coefficient of 1.071 and p-value < 0.0001 for the old voice disguise case. This indicates that, in our material, males were better at disguise than females in both cases.

In the case of prior experience in voice disguise, we do not observe the same effect in old and young voice disguise cases, so that attempting to sound like an old person, was helped significantly by this factor with coefficient 2.33 and p-value < 0.0001. Whereas, trying to sound like a young person, the coefficient is negative -0.803 with p-value 0.002, meaning that

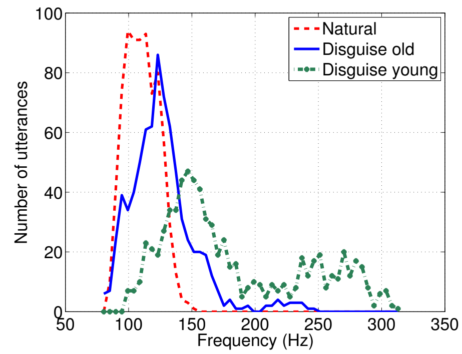


(a) System3 female speakers

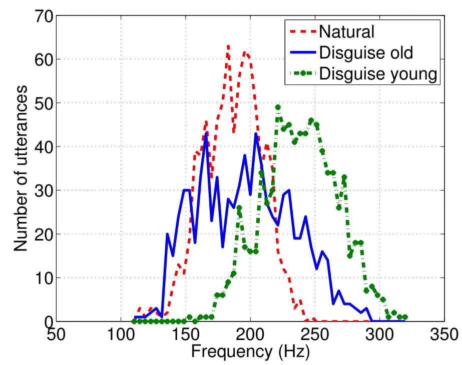


(b) System3 male speakers

Figure 2: DET curves for i-vector PLDA (System3). Results for female and male speakers trials.



(a) Male speakers



(b) Female speakers

Figure 3: Mean $F0$ distributions for the natural and modified voices.

experience in voice disguise actually degraded the speaker's performance to spoof the ASV system in the young voice disguise case.

In the case of smoking habits, it was found that smoking *sometimes* had a degraded effect in disguise ability with coefficient -2.648 and p-value < 0.0001 for old voice disguise, but a small positive effect for young voice disguise coefficient 0.856 and p-value 0.017.

The statistical analysis of our data and the effect of the self-reported information is naturally limited by how well the model fits our data. The significance of the variables were studied independently, so their dependency was not considered in the model outcome. Further, the findings of our analysis are observations on our data only and they should be re-examined on other age disguise corpora.

6. Conclusions

In this paper, we studied the impact of age-related disguise on two widely used ASV methods, GMM-UBM and i-vector. Both systems show a dramatic increase in the equal error rate (EER) when age-related disguise is introduced. For example, for the i-vector PLDA system, the increase of EER is 7-fold for the attempt of old voice and 11-fold for young voice, in the case of male speakers. Similar degradation in performance is observed for female speakers, with a 5-fold EER increase for old voice disguise and 6-fold increase for young voice disguise. The mean $F0$ distributions showed a shift towards higher frequencies when speakers attempted a young voice. This was more or less expected as younger speakers' $F0$ values tend to be higher than for older speakers.

Gender and experience in voice disguise were found to have a significant effect in disguise ability. In the data used for this study, it was found that male speakers were more successful at age-related disguise. For future work, we plan to include a perceptual test to correlate human listening performance and ASV systems in the presence of age-related disguise. The further study of the factors that make age-related disguise successful and its similarities with voice ageing process could shed a light in the further detection of this type of disguise to speaker verification system and further be utilized in automatic disguise detection methods.

7. Acknowledgments

This study was supported by the Academy of Finland (projects no. 253120 and 283256) and the Finnish Scientific Advisory Board for Defense (MATINE) project no. 2500M-003. The authors would like to thank Maria Bentz and Prof. Stefan Werner for their help in planning and executing the data collection project, and to the speakers that produced the needed audio for this study.

8. References

[1] Anil K. Jain, Arun Ross, and Salil Pankanti, "Biometrics: a tool for information security," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 2, pp. 125–143, June 2006.

[2] Anil K. Jain, Arun Ross, and Salil Prabhakar, "An introduction to biometric recognition," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 4–20, 2004.

[3] Andreas Lanitis, "A survey of the effects of aging on biometric identity verification," *International Journal of Biometrics*, vol. 2, no. 1, pp. 34–52, 2009.

[4] Finnian Kelly, *Automatic Recognition of Ageing Speakers*, Ph.D. thesis, Trinity College Dublin, 2014.

[5] Peter Torre III and Jessica A. Barlow, "Age-related changes in acoustic characteristics of adult speech," *Journal of Communication Disorders*, vol. 42, no. 5, pp. 324 – 333, 2009.

[6] Sachin S. Kajarekar, Harry Bratt, Elizabeth Shriberg, and Ray de Leon, "A study of intentional voice modifications for evading automatic speaker recognition," in *Proc. Odyssey*, 2006, pp. 1–6.

[7] Finnian Kelly, Rahim Saeidi, Naomi Harte, and David van Leeuwen, "Effect of long-term ageing on i-vector speaker verification," in *INTERSPEECH*, 2014, pp. 86–90.

[8] George Doddington, "The effect of target/non-target age difference on speaker recognition performance," in *Proc. Odyssey*, Singapore, 2012.

[9] Yun Lei and John H.L. Hansen, "The role of age in factor analysis for speaker recognition," in *INTERSPEECH*, 2009, pp. 2371–2374.

[10] Douglas A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *DSP*, vol. 10, no. 1, pp. 19–41, Jan 2000.

[11] N. Dehak, P.J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *ITASLP*, vol. 19, no. 4, pp. 788–798, May 2011.

[12] Simon J. D. Prince and James H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proc. ICCV*, 2007, pp. 1–8.

[13] University of Helsinki Institute of Behavioural Sciences, "PERSO databases for finnish speech synthesis. PERSO corpus," `http://urn.fi/urn:nbn:fi:lb-2014073053`.

[14] Niko Brümmer and Edward de Villiers, "BOSARIS toolkit," 2011, from `https://sites.google.com/site/bosaristoolkit/`.

[15] N. Brummer, L. Burget, J. Cernocky, O. Glembek, F. Grezl, M. Karafiat, D. A. van Leeuwen, P. Matejka, P. Schwarz, and A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 2072–2084, Sept 2007.

[16] Yen-Liang Shuea, Patricia Keatingb, Chad Vicenikb, and Kristine YuShue, "VoiceSauce: A program for voice analysis," in *Proceedings of the ICPhS XVII*, 2011, pp. 1846–1849.

[17] Paul Boersma and David Weenink, "Praat: doing phonetics by computer [Computer program]," 2015, Version 5.4.09, retrieved 15 June 2015 from `http://www.praat.org/`.

[18] Paul Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetic Sciences*, 1993, vol. 17, pp. 97–110.

A. Text fragments read by the speakers

A.1. The rainbow passage

Sateenkaaritarina
Kun auringonvalo osuu sadepisaroihin ilmassa, ne käyttäytyvät kuin prismat, ja muodostavat sateenkaaren.
Sateenkaari muodostuu valkoisen valon jakaantuessa useiksi kauniiksi väreiksi.
Nämä muodostavat kauniin pitkän kaaren horisontin yläpuolelle päättyen jonnekin sen taakse.
Legendan mukaan sateenkaaren päässä on padallinen sulaa kultaa.
Ihmiset etsivät sitä kuitenkaan mitään löytämättä. Kun joku etsii jotain mahdotonta, sanotaan hänen etsivän kultaa sateenkaaren päästä.

A.2. The north wind and the sun

Pohjantuuli ja aurinko
Pohjantuuli ja aurinko väittelivät kummalla olisi enemmän voimää, kun he samalla näkivät kulkijan, jolla oli yllään lämmin takki.
Silloin he sopivat, että se on voimakkaampi, joka nopeammin saa kulkijan riisumaan takkinsa.
Pohjantuuli alkoi puhaltaa niin että viuhui, mutta mitä kovempaa se puhalsi, sitä tarkemmin kääri mies takin ympärilleen, ja viimein tuuli luopui koko hommasta. Silloin alkoi aurinko loistaa lämpimästi, eikä aikaakaan, niin kulkija riisui manttelinsa.
Niin oli tuulen pakko myöntää, että aurinko oli kuin olikin heistä vahvempi.