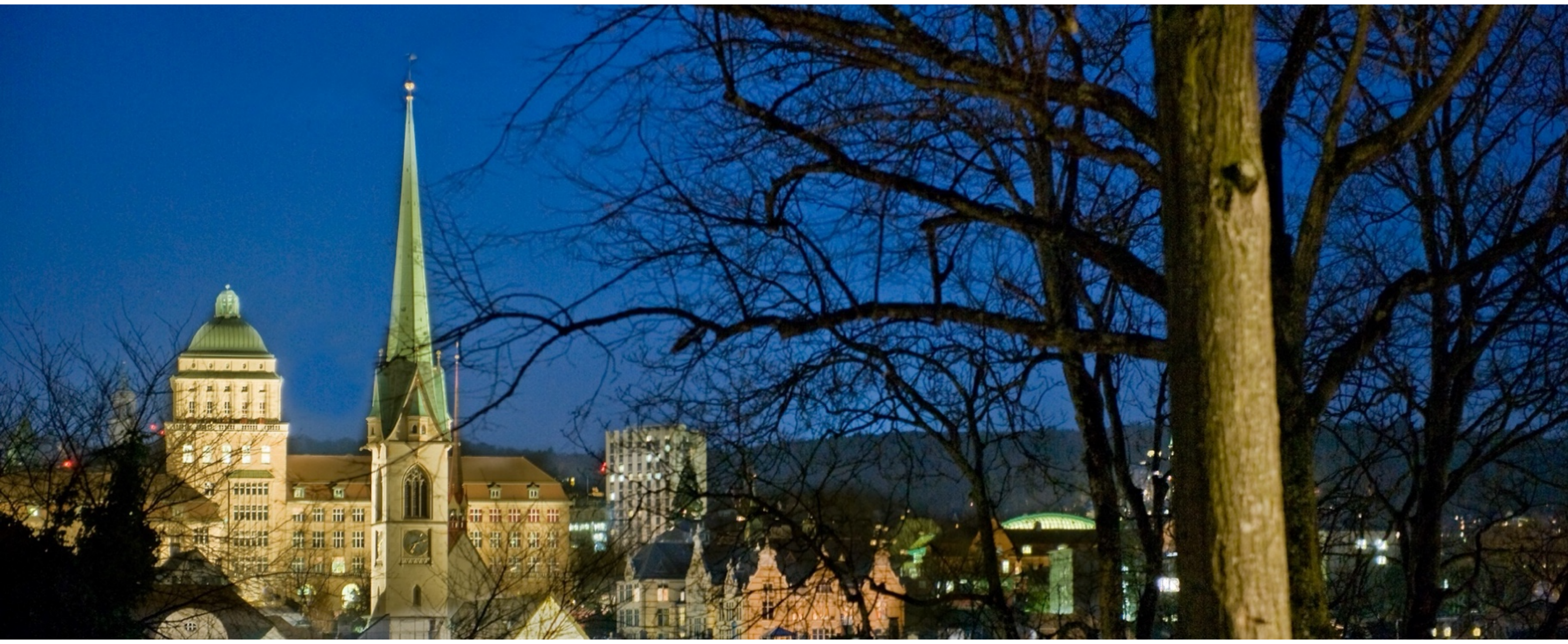
XVII AISV CONFERENCE

Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications

Thursday 4th - Friday 5th February 2021



Book of abstracts

XVII AISV Conference

Associazione Italiana Scienze della Voce

Thursday 4th - Friday 5th February 2021

Hosted by University of Zurich

(online)

Organising Committee

Stephan Schmid (chair),

Camilla Bernardasci, Volker Dellwo,

Dalila Dipino, Davide Garassino, Michele Loporcaro,

Stefano Negrinelli, Elisa Pellegrino,

Dieter Studer-Joho

Student Assistant

Seraina Nadig

Scientific Committee

Table of contents

Forensic value of acoustic-phonetic features from Standard Dutch nasals and fricatives

Laura Smorenburg and Willemijn Heeren
Leiden University Centre for Linguistics

Although vowels generally outperform consonants in speaker discrimination, reports indicate that forensic voice analysts regularly use consonants in auditory-acoustic analysis [1]. However, research on the usefulness of acoustic-phonetic features from consonants in forensic speaker comparisons (FSC) is scarce. We investigated the forensic value of consonants that are highly frequent in Dutch and are therefore likely to be available in forensic material [2]: fricatives (/s x/) and nasals (/n m/). Fricatives are characterised by frication noise at higher or mid-range frequencies, depending on the place of articulation, whereas nasals are characterised by low-frequency energy due to nasal damping. Reports show that place of articulation and uvular trill in the velar/uvular fricative /x/ is strongly associated with region [3] and that sibilant fricative /s/ can carry speaker information such as gender, class, and sexual orientation [e.g. 4, 5]. Subsequent research has shown that /s/ is indeed speaker-specific in Dutch, meaning it has low within and high between-speaker variability [6]. Similarly, nasal consonants exhibit high speaker-specificity because of the nature of a nasal; the involvement of the relatively rigid nasal cavity, which has different shapes and sizes between speakers, results in high between-speaker but low within-speaker variation for nasals [7, p.135]. Because acoustic-phonetic analysis is prevalent in FSC [8], we investigated the forensic value of acoustic-phonetic features from Dutch nasals and fricatives in conversational telephone speech using the statistical framework used in FSC. Based on earlier work on Dutch (nonsense) read speech [6], we hypothesized that /n/ will outperform /m/ and that nasals outperform fricatives in speaker discrimination.

Method

Materials and acoustic analysis. Landline telephone conversations (bandwidth 340-3400 Hz) from adult male speakers of Standard Dutch were analysed [Spoken Dutch Corpus: 9]. From the same 62 speakers, we annotated 3,561 /s/ tokens (per speaker: M = 57, SD = 24), 3,836 /x/ tokens (per speaker: M = 62, SD = 31), 4,676 /n/ tokens (per speaker: M = 74, SD = 28), and 3,654 /m/ tokens (per speaker: M = 58, SD = 24). For fricatives, the following features were extracted per token: duration (log10-transformed), centre of gravity (CoG), standard deviation (SD), skewness (SKW), kurtosis (KUR), and spectral tilt. CoG was also measured in five non-overlapping windows of 20% of a token's duration, after which a cubic polynomial fit was made to capture the dynamics of CoG, resulting in four coefficients. For nasals, we also measured the second and third nasal formants (N2, N3), and their bandwidths (BW2, BW3). N2 and N3 were also captured dynamically, in the same way as CoG.

Statistical analysis. Speaker discriminability was established with likelihood ratios (LR), which reflect the ratio of the probability of the evidence under the hypothesis that two speech samples come from the same speaker (SS) to the probability of the evidence under the hypothesis that two speech samples come from different speakers (DS). The analysis was performed using a MATLAB implementation [10] based on the LR algorithm proposed in [11], where within-speaker variation is modelled as a normal distribution and between-speaker variation is modelled with a multivariate kernel density. LR systems were built for each consonant, using acoustic-phonetic features as parameters. Highly correlating features may inflate the strength of evidence, so a maximum correlation was set at $r = .50$. For /s/ and /x/, this resulted in the following parameters: duration, CoG, SD, Kur, and the three dynamic CoG coefficients. For /n/ and /m/, we used the same parameters for a direct comparison with the fricatives and included the nasal formants and bandwidths in a separate system.

Per system, the 62 speakers were divided into a development (N=22), reference (N=20), and test set (N=20). First, SS and DS LRs were computed for the development set. Not all speakers had multiple recordings, so the tokens per speaker were divided in half to generate SS comparisons. For the development set, this resulted in 22 SS and 231 DS comparisons. The LR scores from these comparisons were used to obtain calibration parameters (shift, slope) for the test set. LLRs were then obtained and calibrated for the test set. To reduce sampling effects, 10 iterations were used in which the development, reference, and test sets were sampled at random. The systems' performance was assessed through SS and DS LLRs and the log-likelihood-ratio costs (C_{llr}), which reflects the degree of accuracy of the system's calibrated decisions. Median LLRs and C_{llr}s over iterations were obtained using R package *sretools* [12].

Results

Table I displays the results. An LLR of 1 means that the evidence is 10 times more likely under the same-speaker (SS) hypothesis and an LLR of –1 means it is 10 times more likely under the different-speaker (DS) hypothesis. E.g., the LLR_{SS} of 1.52 means that the evidence is 33 times more likely under the SS hypothesis than the DS hypothesis. For C_{llr}, closer to 0 is better.

Table I. Median SS and DS LLRs and C_{llr}s

	Static parameters			Dynamic parameters			Static nasal-specific parameters			Dynamic nasal-specific parameters		
	LLR_{SS}	LLR_{DS}	C_{llr}	LLR_{SS}	LLR_{DS}	C_{llr}	LLR_{SS}	LLR_{DS}	C_{llr}	LLR_{SS}	LLR_{DS}	C_{llr}
/s/	1.52	–2.36	0.52	0.25	–0.10	0.91						
/x/	0.74	–0.20	0.82	0.26	–0.03	0.96						
/n/	0.74	–0.60	0.67	0.43	–0.08	0.87	1.55	–1.54	0.55	0.13	–0.08	0.96
/m/	0.85	–0.50	0.71	0.21	–0.07	0.93	1.05	–0.78	0.70	0.03	0.01	0.99

Discussion and conclusion

Results indicate that /s x n m/ have forensic value, but that the extracted acoustic-phonetic features differ in their discriminatory power. Static acoustic-phonetic features contained more speaker information than dynamic acoustic-phonetic features. This is perhaps due to contextual influences in these short consonants leaving little speaker-specific information in the dynamics. Nasals performed better with static nasal-specific features. Against expectations, we found that /s/ outperformed the other consonants, even though it was sampled from telephone speech and its spectral peak falls outside of the telephone band.

Acknowledgement NWO VIDI grant (276-75-010) supported this work.

References

[1] Gold, E., & French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law, 18*(2), 293–307.

[2] Luyckx, K., Kloots, H., Coussé, E., & Gillis, S. (2007). Klankfrequenties in het Nederlands. In *Tussen taal, spelling en onderwijs* (pp. 141–154). Academia Press.

[3] Harst, S. Van der, Velde, H. Van de, & Schouten, B. (2007). Acoustic characteristics of Standard Dutch /x/. *Proceedings of the 16th ICPhS*, 1469–1472.

[4] Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *J.Phon., 34*, 202–240.

[5] Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. *Change in Phonology: Papers in Laboratory Phonology, 9*, 65–86.

[6] Van den Heuvel, H. (1996). *Speaker variability in acoustic properties of Dutch phoneme realisations*, Radboud Universiteit, Nijmegen.

[7] Rose, P. (2002). Forensic Speaker Identification. In *Sciences New York* (Vol. 20025246).

[8] Gold, E., & French, P. (2019). International practices in forensic speaker comparisons: Second survey. *International Journal of Speech, Language and the Law*, *26*(1), 1–20.

[9] Oostdijk, N. H. J. (2000). Corpus Gesproken Nederlands. *Nederlandse Taalkunde, 5*, 280–284.

[10] Morrison, G.S. (2007). Matlab implementation of Aitken & Lucy's (2004) forensic likelihood-ratio software using multivariate-kernel-density estimation. [software].

[11] Aitken, C. G. G., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *J. of the Royal Stat. Soc. Series C: Applied Statistics*, *53*(1), 109–122.

[12] Van Leeuwen, D. (2011). SREtools: Compute performance measures for speaker recognition.

Universität
Zürich[UZH]