Speaker Verification Based on the German VeriDat Database

Ulrich Türk, Florian Schiel
{tuerk,schiel}@bas.uni-muenchen.de
Bavarian Archive for Speech Signals (BAS)
c/o Institute of Phonetics and Speech Communication
University of Munich

Abstract

Here we introduce

- the new German speaker verification (SV) database VeriDat of telephone speech,
- the baseline performance of our experimental speaker verification system
- and several experiments with different data splittings according to the network type and the acoustical environment.

Motivation

"Real world" telephone data show several difficulties, namely

- varying acoustical environments and telephone channels;
- linguistic and meta-linguistic "noise";
- erratic speaking style for some speakers ("goats" have unusually high false-rejection rate);
- sparse material for building models.

⇒ Introducing some knowledge about certain properties of the data might give a better performance in the verification task.

⇒ Further analyzing of the data might give clues to goat-like behaviour and how it could be detected a priori.

The VeriDat database

- extension of standardized specification for SV databases in the SpeechDat project.
- 150 speakers covering the whole range of German dialects.
- Speech items used here: number triples (e.g. "21 35 76").
- different networks: fixed and cellular
- different background noises: quiet and noisy.

System Design

Setup & Experimental Protocol

speaker set	# of speakers
client	30
world	30
impostor	60
development	30

Features, Modelling and Score Normalization

- standard HTK features (LPCC).
- word based HMMs, left-right topology.
- temporal segmentation and normalization with world model.
- EERs are gender-balanced using a posteriori speaker-dependent thresholds.

Experiments

- same amount of training data for world and client models.
- cheating experiments in the domain "Fixed/GSM" (F/G) and "Quiet/Noisy" (Q/N).

Exp.	World Model(s)	Client Model(s)
Base	one model: all cond. 1260 recs.	one model: all cond., 20 recs.
FQ	one model: FixedQuiet 1260 recs.	one model: FixedQuiet 20 recs.
F/G 2/2	two models: Fixed, 1260 recs. GSM, 1260 recs.	two models: Fixed, 20 recs. GSM, 20 recs.
F/G 1/2	one model	two models
F/G 2/1	two models	one model
Q/N 2/2	two models: Quiet, 1260 recs. Noisy, 1260 recs.	two models: Quiet, 20 recs. Noisy, 20 recs.
Q/N 1/2	one model	two models
Q/N 2/1	two models	one model

Test parameters for base line and cheating experiments.

Results

Experiment	EER mean %
Base	3.93
FQ	2.30

Results for the baseline experiments.

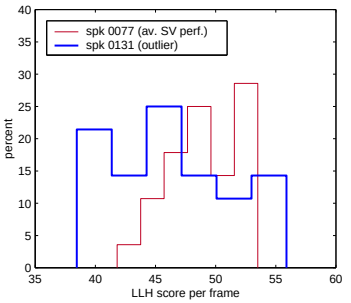
# World Models	# Client Models 1	2
1	3.93	4.45
2	3.89	4.22

Results (EER mean %) for the cheating experiment F/G.

# World Models	# Client Models 1	2
1	3.93	3.66
2	4.47	3.93

Results (EER mean %) for the cheating experiments Q/N.

"Goat-like" behaviour



Histogram of LLH scores per frame of the world model using client's enrollment data.

Conclusion

- Base line performance deteriorates when including all acoustical variations in the training material.
- Splitting the training material according to the network ("Fixed/Quiet") does not result in any improvement.
- Splitting the training material according to the background noise ("Quiet/Noisy") gives a slight gain in performance.
- A few client speakers show extremly low verification performance.
- Detection of "goat-like" behaviour is a crucial point for a real world SV system.

Online-Resource

http://www.bas.uni-muenchen.de/Bas/SV/

Eurospeech 2003 – Geneva, Switzerland; Poster Session "Speaker Recognition & Verification"; Thursday, September 4, 13:30–15:30