NFI-FRITS: A forensic speaker recognition database

David van der Vloed[1], *Jos Bouten*[2]
[1]*Netherlands Forensic Institute*
d.van.der.vloed@nfi.minvenj.nl
[2]*visiting scientist*
We prefer a poster presentation

The NFI-FRITS database (Forensically Realistic Intercepted Telephone Speech) contains speech intercepted during real police investigations. This material was obtained to facilitate research on data typically encountered in forensic practice, much like the data used by Becker (2012) and Van Leeuwen and Bouten (2004) and the AHUMADA III data (Ramos et al, 2008). NFI-FRITS consists of over 4100 recordings of more than 600 speakers.

Data processing

The raw data were provided with some metadata, like the two telephone numbers involved in the telephone call, case name, etc. The audio files were split in two single channel files (a and b) and stored in a database, along with the provided metadata.
Native listeners listened to the material and removed information in the audiofiles that can identify an individual and assigned speaker names and other metadata. This was done until about five recordings were assigned to a speaker, after which the process was repeated.

Realistic data

The database consists of realistic data, meaning that the audio comes from intercepted telephone speech from police investigations. The forensic nature of the recordings and the method to label a recording with a speaker name make the truth about speaker identities in this database a truth by proxy. Nevertheless the authors feel that the method used leads to sufficiently reliable speaker identities. The data is representative of police investigations, however, the collection is not representative of casework at the NFI as this typically involves recordings where the speaker ID is disputed.

Database by numbers

The database consists of 4188 recordings and 604 speakers. There are 427 male speakers in 3120 recordings and 177 female speakers in 1068 recordings. There are 72 multilingual speakers in the database, who speak Dutch in some recordings and either Turkish, Moroccan Arabic or Berber in other recordings.

Table 1. No. recordings per language

Language	#recordings
Dutch (all varieties)	3091
Turkish	499
Moroccan Arabic	191
Berber (Tarifit)	116
Mixed	245
Other	46

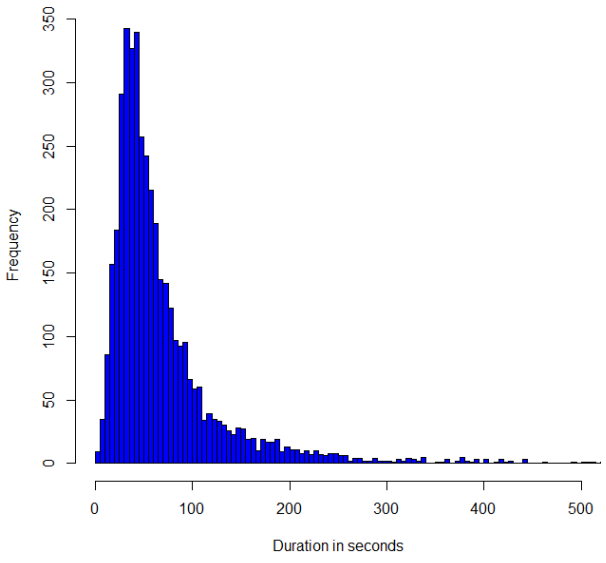

Figure 1 Histogram of measured durations of speech per recording

References

T. Becker, “Automatic forensic voice comparison (automatischer forensischer stimmenvergleich),” in The Journal of Speech Language and the Law, 2012, vol. 19, pp. 291–294.

David A. van Leeuwen and Jos S. Bouten, “Results of the 2003 NFI-TNO forensic speaker recognition evaluation,” in Proc. Odyssey 2004 Speaker and Language recognition workshop. June 2004, pp. 75–82, ISCA.

Daniel Ramos, Joaquin Gonzalez-Rodriguez, Javier Gonzalez-Dominguez, and Jose Juan Lucena-Molina, “Addressing database mismatch in forensic speaker recognition with Ahumada III: a public real-casework database in Spanish.,” in Proc. Interspeech, 2008, pp. 1493–1496.