Biometric Speaker Recognition Evidence within a Court Environment

Hermann J. Künzel
Dept. of Phonetics, University of Marburg, Marburg, Germany
kuenzelh@staff.uni-marburg.de

Biographic Basics

- M.A. and Ph.D. in Phonetics
- Head of Speaker Identification & Tape Authentication Dept. of the German BKA from 1985 – 1999
- Professor of Phonetics, University of Marburg, since 1999
- International expert on forensic SPID for law enforcement authorities, the judiciary and private customers since 1980

Main tasks
in forensic speech processing

Speaker identification (SPID)

- by witnesses and victims (voice line up)
- by experts (speech scientists)
  - (a) Classical phonetic-acoustic method (2-vector approach)
  - (b) Automatic speaker identification by computer
  - (c) Combination of (a) and (b) (3-vector or “hybrid approach“).

Main tasks
in forensic speech processing

Disputed utterances

decoding & transcription of distorted speech

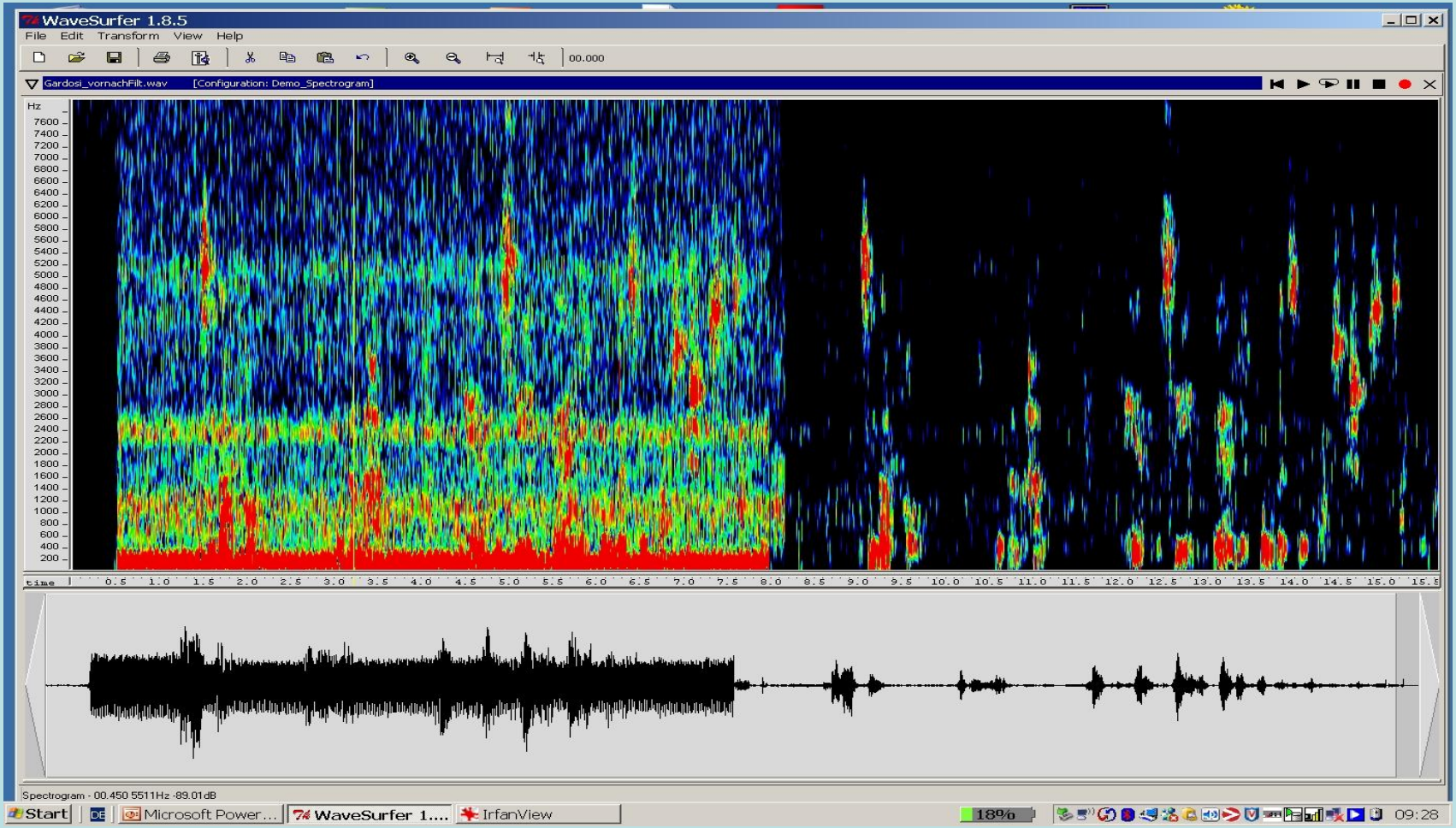
Speech enhancement

improving acoustic quality for listening & technical analyses using digital signal processing tools

Frequently, both tasks arise. Example (sexual assault, rape):

Speech signal before (left side) and after enhancement. Oscillogram and three-dimensional spectrogram of the utterance

“Sie war faul, und dann hat er sie während der Probezeit rausgeschmissen, sagt er.“ (“She was lazy and then he fired her during the probationary period he says“)



Peculiarities of forensic speech recordings

1. Non - cooperative speaker

Individual has no interest in being identified; may disguise voice and/ or speech (15 - 50 % of cases, depending on type of offence)

2. No control on type and amount of speech material

No word-by-word comparisons possible in most cases; interruptions of speaking process by other speaker(s)

3. Telephone transmission of speech signal

(ca. 95% of cases) causes limitations of frequency and S/N ratio, distortions by mains hum, stochastic impulses, GSM codecs etc.

4. Open set of potential speakers

virtually infinite, e.g. all adult males with Czech as first-language.

(a) The classical approach: Acoustic-phonetic SPID

Acoustic phonetic SPID

Step 1:

Diagnosis & extraction of phonetic, acoustic & medical features of voice, speech and language (characteristics of pitch, voice quality (e.g. hoarseness), dialect, jargon, breathing behaviour etc.)

Step 2:

Assessment of the speaker-specific (idiotypical) power of these features (medical statistics; "private" statistics based on expert‘s (and experts‘) individual experience)

Principles of Acoustic – Phonetic SPID

Step 3:

Synoptical interpretation of results for all parameters determines the identity statement made on an *impressionistic* probability scale

Principles of Acoustic – Phonetic SPID

Step 3:

Synoptical interpretation of results for all parameters determines the identity statement made on an impressionistic probability scale

in Germany and many other jurisdictions:

„Identity / non-identity

is possible / probable / very probable / beyond reasonable doubt"

in UK:

“Identity is consistent / not consistent with the data / no decision“

Two-tier analysis of speaker-specific features

Let‘s go back to Step 1 for a minute:

Diagnosis & extraction of linguistic, acoustic & medical features of voice, speech and language (characteristics of pitch, voice quality (e.g. hoarseness), dialect, jargon, breathing behaviour etc.)

Highlight of the method originally developed by the BKA (Künzel 1987) is the **two-tier analysis** of all features:

(1) phonetic, linguistic, medical level (mainly auditive)

(2) acoustic, physical, statistic level, providing an **objective assessment of speaker-specific features.**

Here are some tools for (2):

Two-tier analysis of speaker-specific features

Main tools for the objective documentation of subjective perceptions:

- Oscillogram (display of sound pressure as a function of time)
- RMS curve (display of acoustic power as a function of time)
- Automatic determination of voice pitch Fo (ave., variation)
- 2-dimensional spectrogram
- 3-dimensional spectrogram (sonagram)

Here is an illustration:

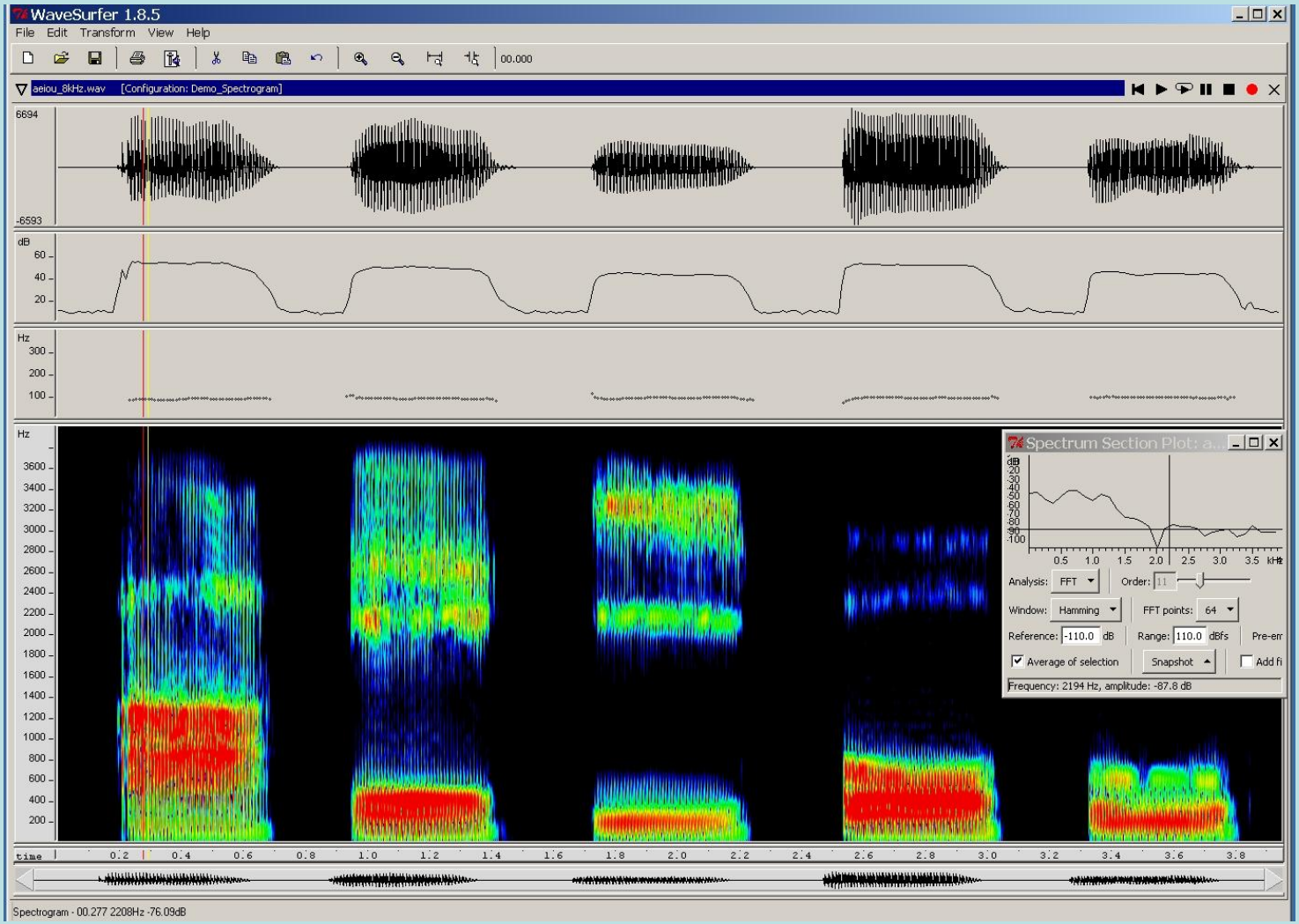
Acoustic documentation of phonetic features

oscillogram

power plot

voice pitch

3-dim. spectrum

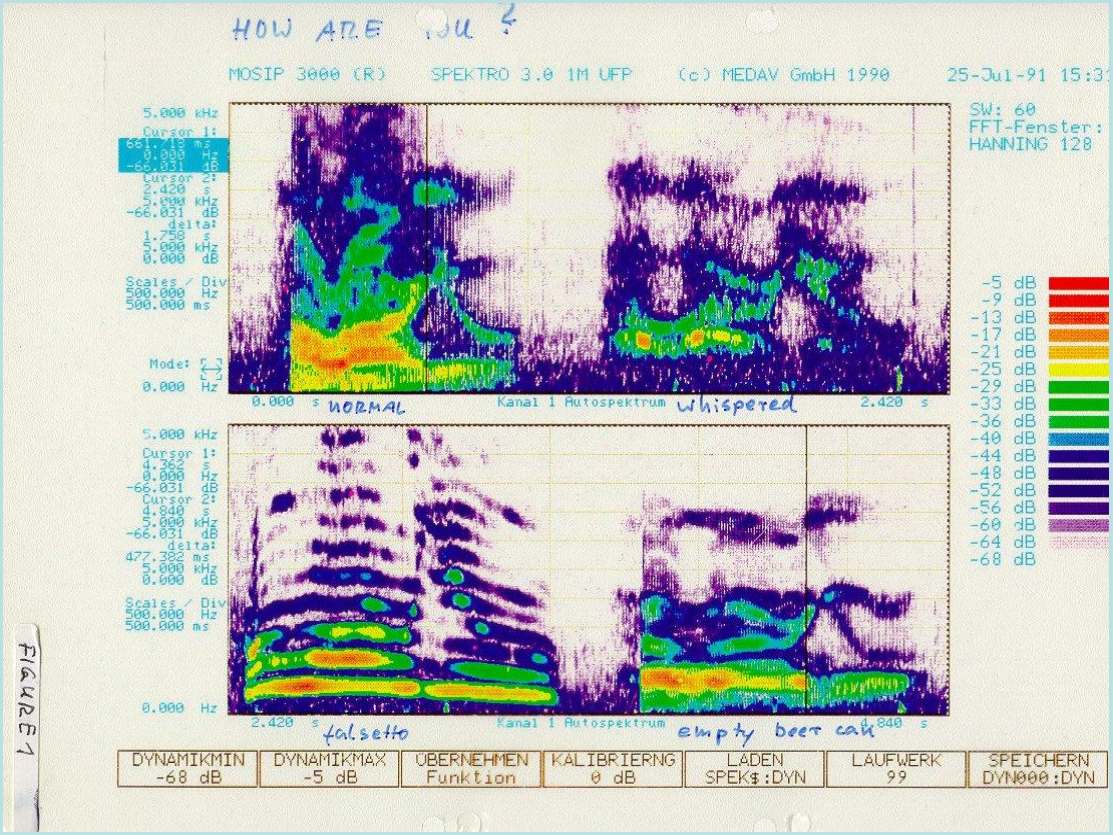


2-dim spectrum

a e i o u

Acoustic documentation of phonetic features

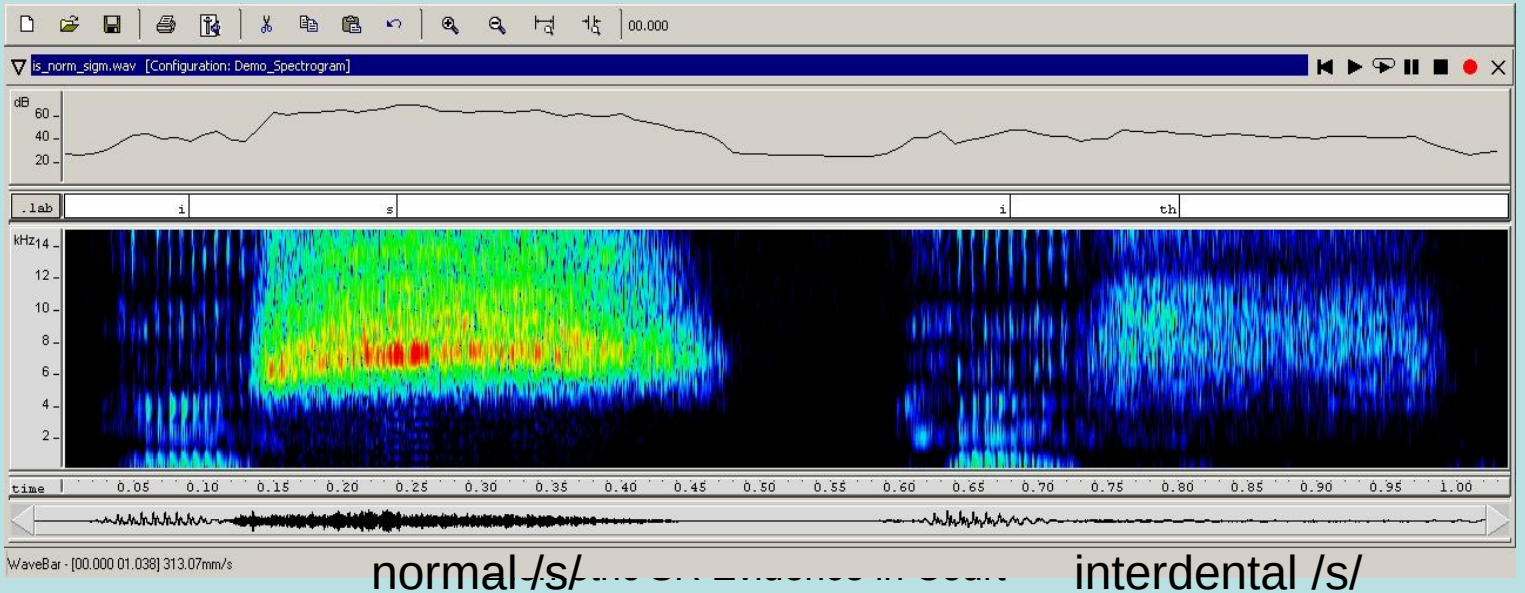
Acoustic effects of voice disguise



Two-tier analysis of speaker-specific features

Example 1

- mispronunciation of consonant /s/ detected auditively (sigmatismus interdentalis partialis)
- diagnosis supported, and quantified in terms of frequency and amplitude by spectral analysis
- Medical statistics reveal 5 per cent of adult [German] males share mispronunciation of /s/

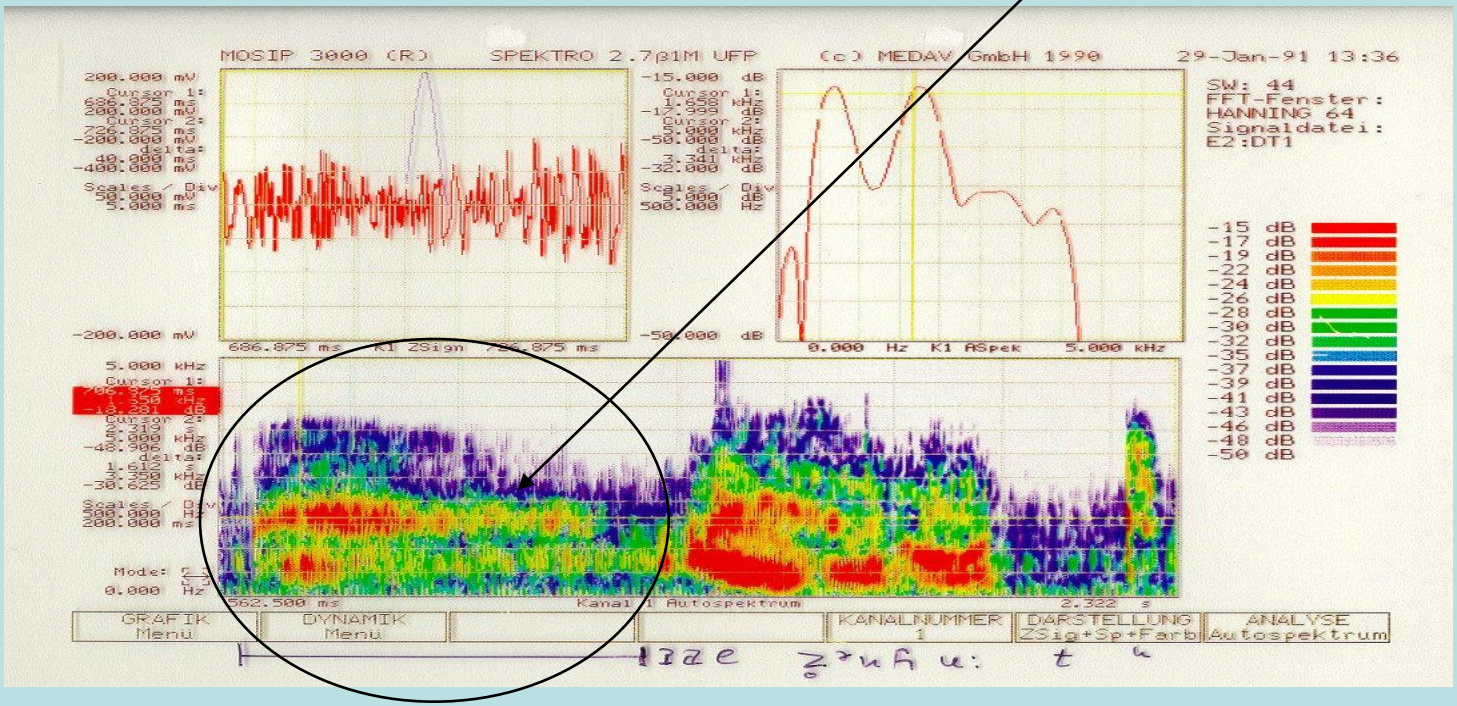


normal /s/ interdental /s/

Two-tier analysis of speaker-specific features

Example 2

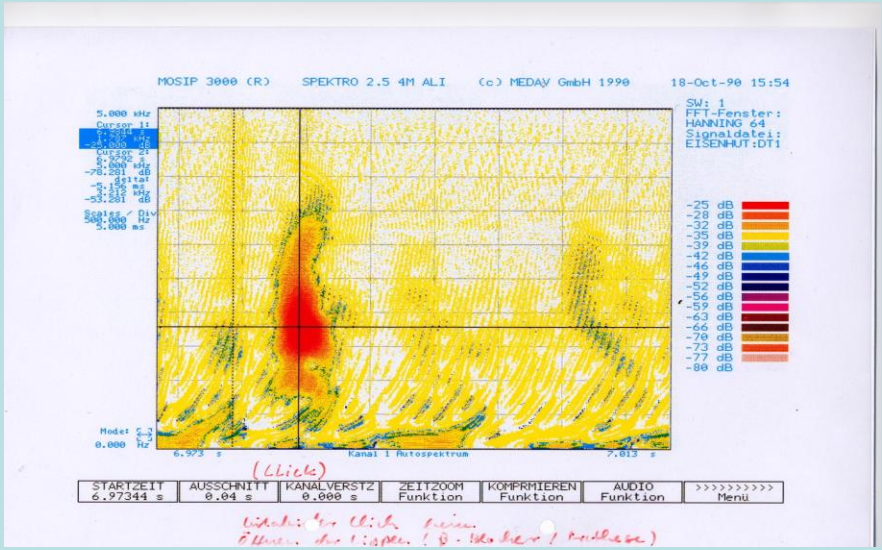
- Pathological breathing detected auditively,
- Acoustic verification using spectral analysis (2 formants!)
- Later confirmed as symptom of hayfever:



Two-tier analysis of speaker-specific features

Example 3

- Click sound discovered auditively when speaker opens lips at start of message
- Acoustically confirmed using spectral & power analysis
- Physiological explanation: alteration of viscosity of saliva
- Known causes: non-perfect prosthetic devices, certain drugs (e.g. beta blockers)



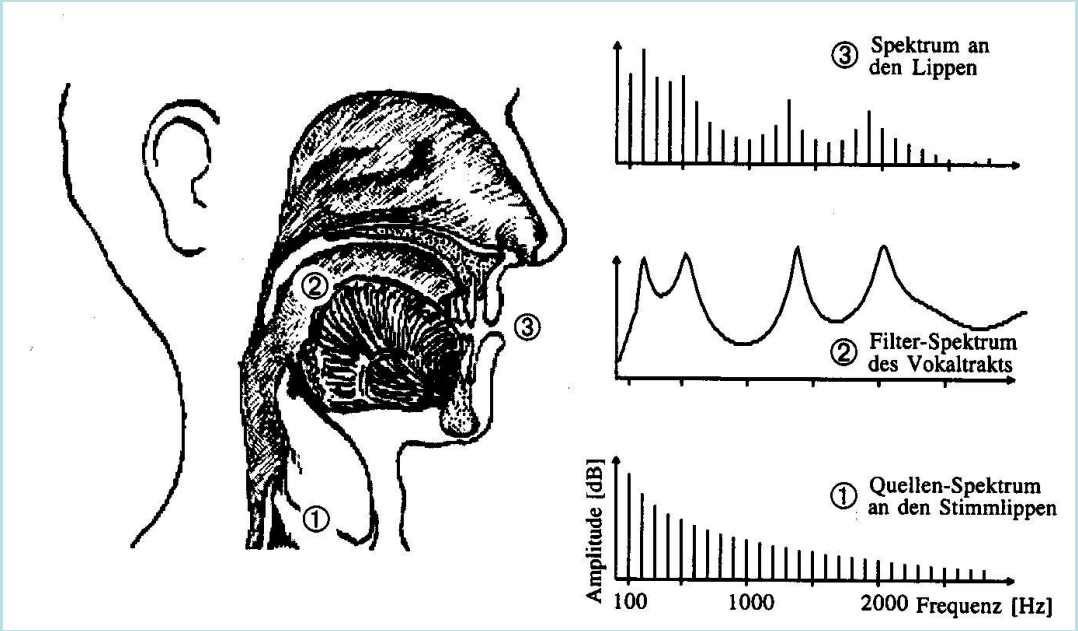
The snags of the acoustic–phonetic SPID

- Unlike fingerprints or DNA, features of voice and speech are not invariable individual traits
- The individualising power of voice and speech features may be affected by the forensic environment (transmission channel characteristics, voice disguise, etc.) [Applies also to the parameters of Automatic SPID!]
- Significant human factor involved (scientific background / professional experience / skills of the expert)
- Impressionistic and subjective verbal probability scale.

(b) The novel approach: Automatic Systems for SPID

Features of automatic systems for forensic SPID

- Acoustic parameters (mostly set of Mel FCCs) map features of the individual resonance behaviour of a speaker‘s vocal tract, i.e. the anatomical and physiological determinants of the mouth, nose and pharynx.



Spectrum at the lips

Filter spectrum generated by vocal cavities

Spectrum at the vocal folds

Features of automatic systems for forensic SPID

- This principle is the basis for the **most powerful of all features** because it implies **independence of language** (so-called *total-voice* system)
- Put differently: It means that the expert no longer needs to know the language(s) involved

- Excursion: A typical case

Automatic SPID: a typical case

- International gang of criminals steal luxury vehicles in Europe, ship to Middle East
- Police collect > 6.000 telephone calls
- Voice samples from a dozen suspects available
- Suspects speak mainly Indian and Pakistani languages, some speak more than one language
- Police face problem to find (enough) interpreters to transcribe the calls (quickly enough) to keep track of the criminals.

Imagine the time and money needed to transcribe 6.000 calls - even if they did have enough translation power!

Automatic SPID: a typical case

- Digital speech files from wiretaps and suspects are presented to Speech Crime Lab.
- Request:

  “Estimate with a reasonable degree of certainty which suspect is a speaker in any of the intercepted calls.“

- Results are presented within a few weeks, streamlining and speeding up investigations. Case is eventially solved.

Automatic SPID: Voice Spotting

Transfer this setting to applications in the intelligence environment, in particular to

- Voice Spotting

  i.e. surveillance of a (usually large) number of communication channels for a set of known voices (targets), such as terrorists.

Voice Spotting: sequence of events

1. Pre-set similarity threshold between known and unknown voices is met
2. Automatic storage of the respective conversation is initiated and a human monitor is alerted.
3. Full-scale voice comparison, preferably of the hybrid type presented below, will further clarify the identity question (requires ca. 1 day).
4. Depending upon computing power, nos. of channels and nos. of target voices, the automatic process (of voice spotting) takes any time between a minute and a day.

More features of automatic systems for SPID

- Powerful algorithms mitigate the problem of unknown channel characteristics ("channel normalisation")
- Decisions on identity of unknown and suspect voices are based on Bayes' theorem (like DNA analysis), results are typically presented as likelihood ratios (LR; prosecutor‘s hyp. / defense hyp.).
- Two major advantages:
  - Numerical scale is superior to subjective verbal (ranking) scales.
  - Results on voice evidence can be combined (multiplied) with results of other evidence assessed in the same way, e.g. imprints of shoe soles, car glass fragments or tyre imprints.

A hybrid approach to SPID:

Please consider:

Parameters used with the classical acoustic phonetic approach and those used by the automatic system are almost entirely **statistically independent.**

Therefore, an obvious question arises:

Why not join the strongholds of both methods into one forensic report ?

Such a “hybrid approach" is currently regarded as the gold standard in forensic (!) SPID, particularly in a report for the Court.

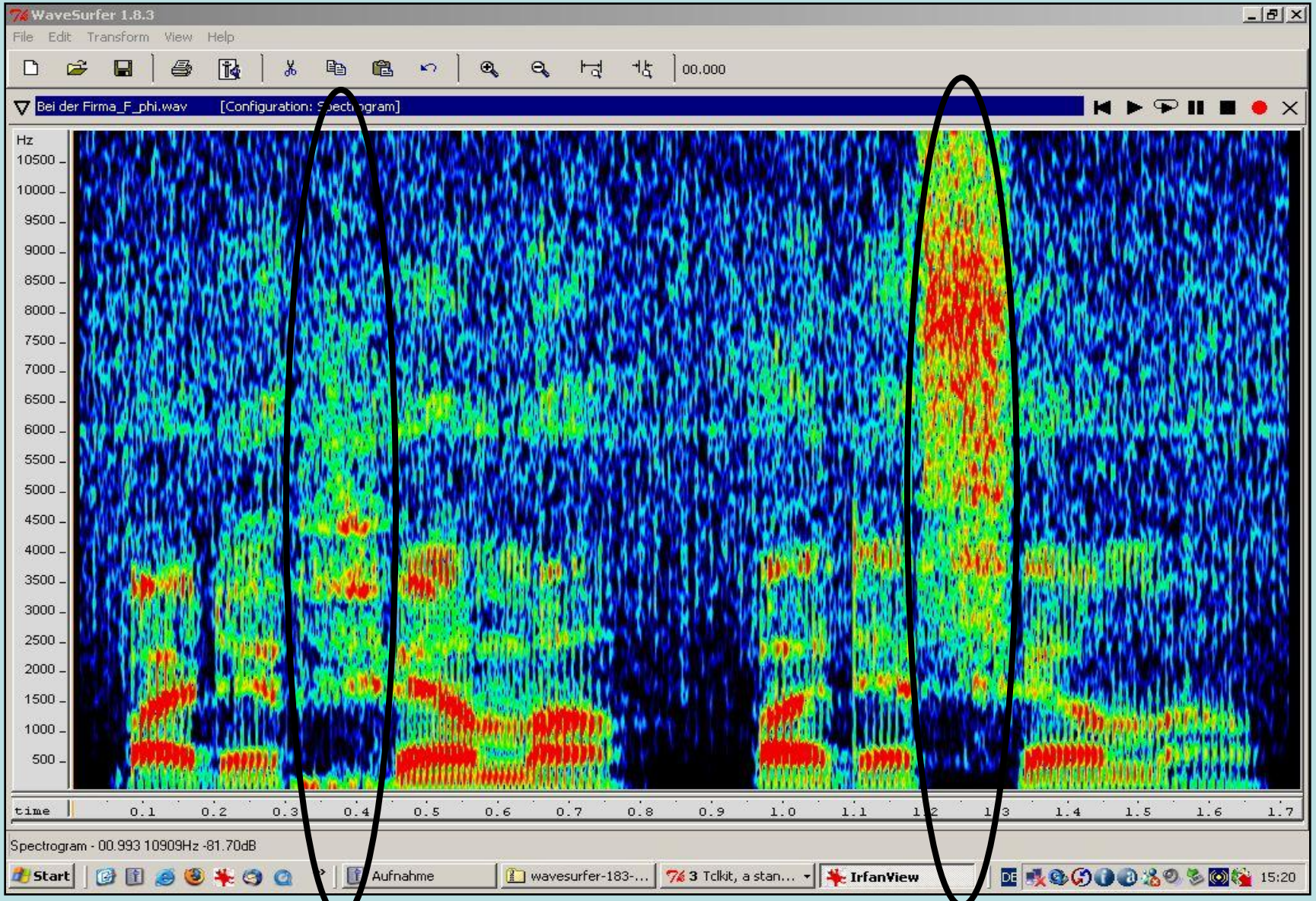
Demonstration of the hybrid approach in two typical forensic cases

Case 1: The facts

- Male speaker
- 2 hoax bomb calls to automotive supplies plant, causing evacuation of > 2,000 employees (financial loss ~ 200,000 €)
- Voice was heavily disguised, total speech ca. 20 secs.
- Taped calls are played to some staff
- 1 suspect (50) vaguely „identified“ by 2 workmates; denies all allegations
- State Prosecutor applies for voice identification report

Case 1: The phonetic side

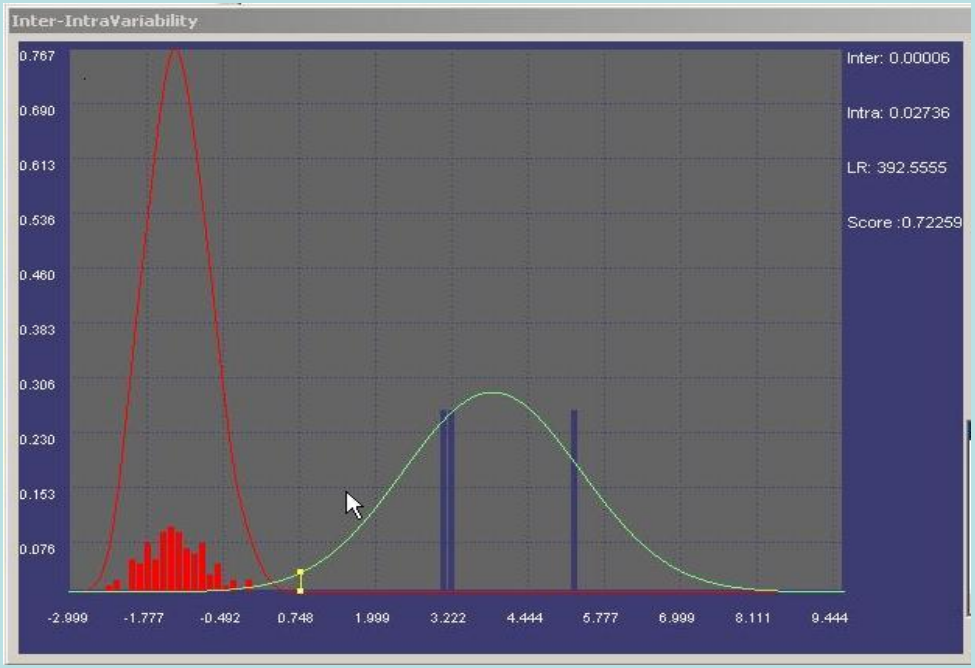
- Most salient feature in phonetic and acoustic analysis: absence of spectral energy above 2.5 kHz in both speakers‘ speech sound [f], resulting in perfect bilabial [Φ] (see next slide)
- Cause emerges when suspect produces reference sample: all 4 upper incisors missing, no prosthesis worn!
- All other parameters very similar / identical, esp. distinct regional colouring



teethless [f]

normal [f]

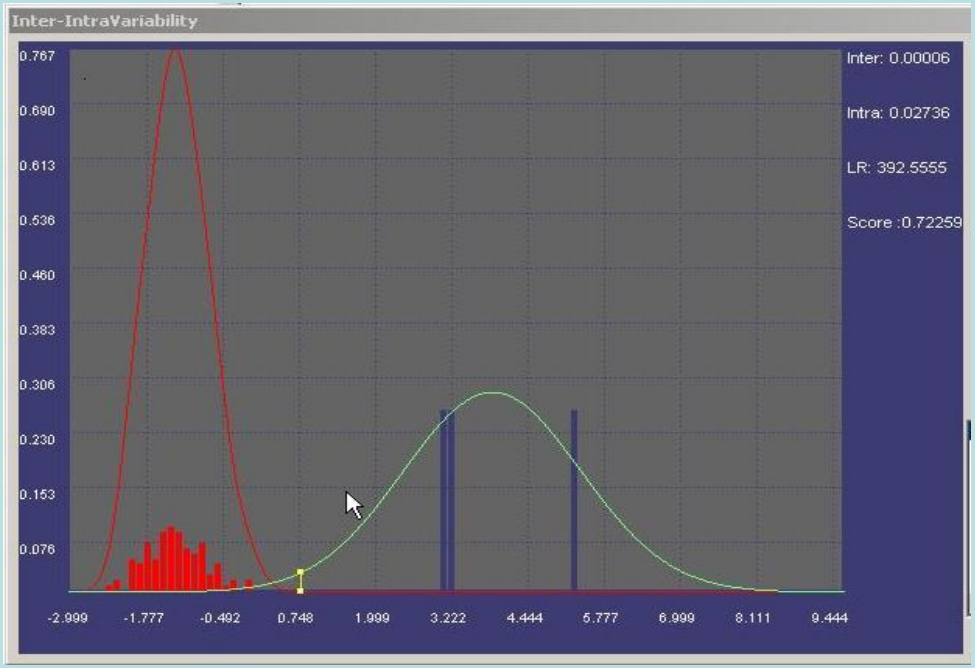
Case 1: Output of Automatic System (Batvox 1)



Red = reference pop; green = intra-suspect distribution

Yellow bar = questioned call (bomb threat)

Case 1: Output of Automatic System (Batvox 1)



- ✓ Identity is 392 times more likely than non-identity.
- ✓ Good result considering the voice disguise!

Case 2: The facts

- Serial fraud case (> 200 cases)
- Modus operandi: “the grandson trick“
- Several male speakers, all close relatives
- Set of 123 calls (Polish, Sinto, German) pre-selected by the Court from > 2,000 calls (ranging from 20 s to several minutes)
- 1 Defendant (48 yrs), multilingual Gypsy (Polish, Sinto, German)

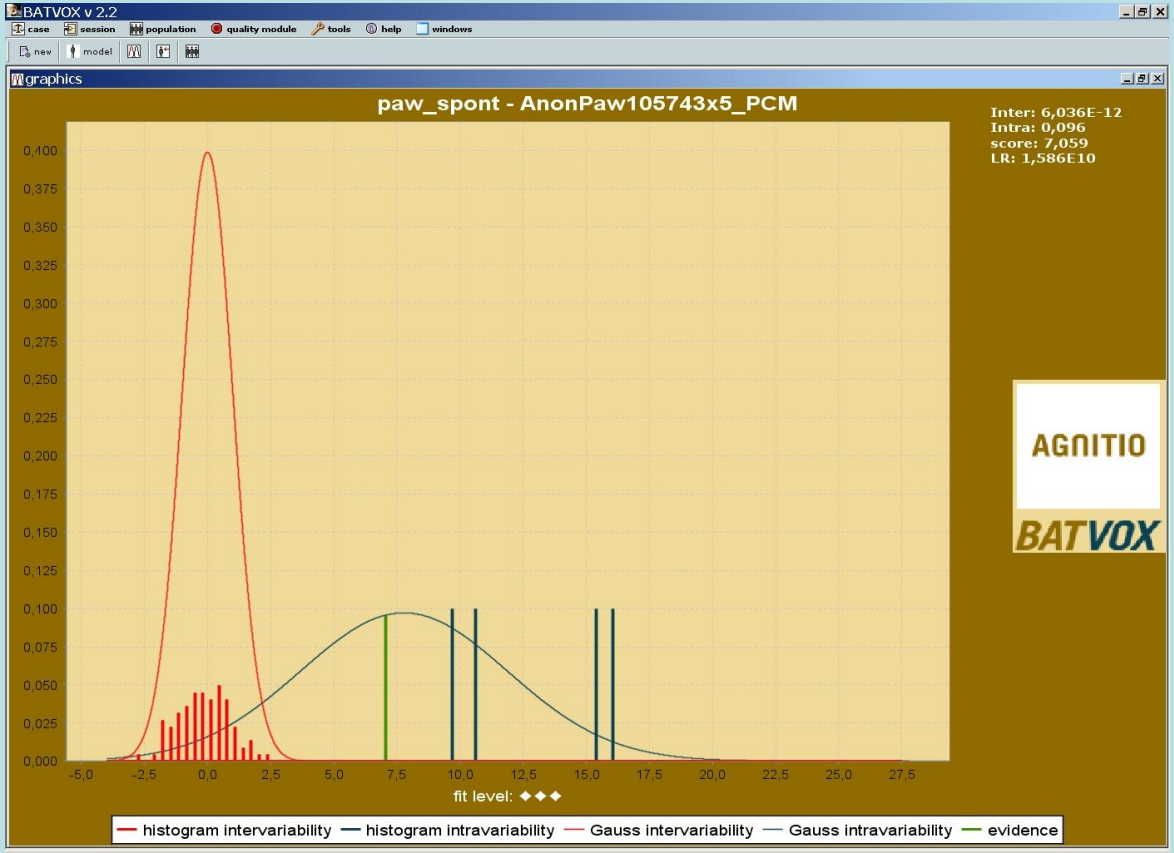
Case 2: The phonetic side

- Characteristic voice quality (hoarseness, high jitter index)
- Difficulties to find speech-related features for comparisons due to three different languages involved
- Conclusion for 122 of the 123 calls.

  1. Phonetic-acoustic method: Probability of identity of suspect and Anonymus varying from "possible" to "very high", depending on amount of material (duration of resp. call)

  2. Automatic SPID: LRs between 1,000 and >> 1,000,000
- Conclusion for one call: "very high" probability of non-identity of suspect and Anonymus / LR below 1,0.

Case 2: The automatic side. Output of Batvox 2



- In court, defendant admits to being the anonymous talker in 122 calls immediately after presentation of report (including this graph)

Case 2. Please consider:

- Given the total of 123 calls (with more than 200 male speakers) it would have been virtually impossible to solve the case using *only* the acoustic-phonetic method (time, cost, legal deadlines)
- Three different languages are involved, which is a severe handicap to the phonetic method, since the number of comparable parameters is reduced, which entails a weaker statement on identity.

General conclusions (1)

1. Automatic methods for forensic speaker identification have become a powerful complement to the classical approach.
2. Cases involving large amounts of, and / or, multilingual speech data may be impossible to solve without the help of an automatic method
3. Provided that prerequisites such as an appropriate model for the reference population (“world population“) are available, automatic systems take only minutes to deliver a result.
4. If an automatic SPID system has passed its fire test in the extremely hostile forensic environment there should be no problem to adapt it to much more benign "commercial" applications (e.g. access control, E – banking).

General conclusions (2)

5. If voice ID reports have to presented as evidence in court the traditional acoustic phonetic method is of special importance: The n*atural speech* parameters such as high/ low voice pitch, monotonous intonation, dialect, foreign accent, speech defects etc. are easy to explain and demonstrate to the jury / judge / opposing expert. On the other hand, the parameters of automatic systems are abstract and difficult to explain to lay people.
6. Currently the best method is a combination of acoustic-phonetic and total-voice automatic approaches.

General conclusions (3)

7. Considering typical applications of SPID in the intelligence environment where, at a certain time, speedy results may be more important than highest certainty, the performance of automatic systems is unbeatable in providing facts to decision makers.
8. If the results have to be used in court trials additional coustic-phonetic expertise is required in order to comply with the current state of the art (hybrid method“).

If you need information on …

- Technical equipment or
- Numbers and professional qualifications of staff needed to operate a speech crime lab
- Literature on phonetic and automatic forensic speaker identification

… feel free to meet me after this talk.

Thank you for your attention !

kuenzelh @ staff.uni-marburg.de