Bundeskriminalamt

Forensischer Stimmenvergleich: Der Likelihood-Ratio-Ansatz

16.01.2013, Universität Wuppertal

Vortragsreihe *Komplexität der Sprache*, ZefiS

(alle Folien auf Englisch)

Michael Jessen
BKA, Fachbereich für Sprechererkennung, Tonträgerauswertung und Autorenerkennung (KT54)

Michael.Jessen@bka.bund.de

Structure

1. Forensic voice comparison: Tasks, methods, illustrations
2. The Likelihood-Ratio approach: principles
3. LR-approach applied to automatic speaker recognition
4. LR-approach applied to Long-Term Formant analysis (LTF)
5. Conclusions

Forensic Voice Comparison

Situation: There is an audio recording of a „questioned speaker" (i.e. the person who committed a crime) and of a suspect. The police/court wants to know: is the questioned speaker identical with the suspect or is the questioned speaker someone else?

Task: Analyze the audio recording and provide a probabilistic statement that is relevant to this question.

Illustrations: Questioned (criminal) Suspect

Drug dealing

Child kidnapping

Extortion

Terrorism

(Die acht Audiodateien, die im Vortrag vorgespielt wurden, können leider nicht öffentlich gemacht werden)

Matching conditions

Mismatched conditions

Methods of voice comparison analysis

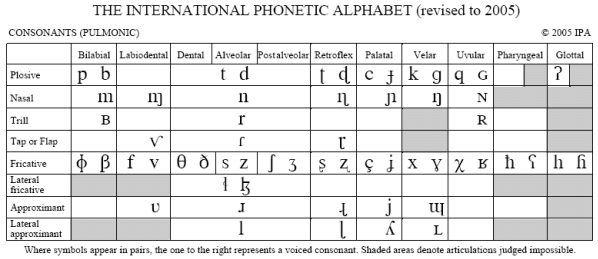
1. auditory-phonetic and linguistic analysis (regional/social varieties and „idiolect"; „paralinguistic" features, such as voice quality, fluency interruptions, breathing patterns, speech pathology)

2. acoustic-phonetic analysis (e.g. f0, formants, articulation rate)

auditory-acoustic approach

3. Automatic speaker recognition

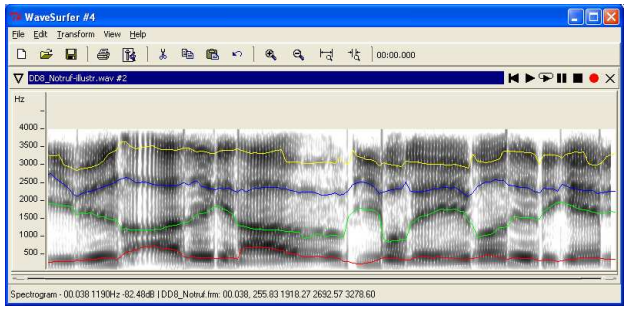
Jessen (2012) for more details


THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.



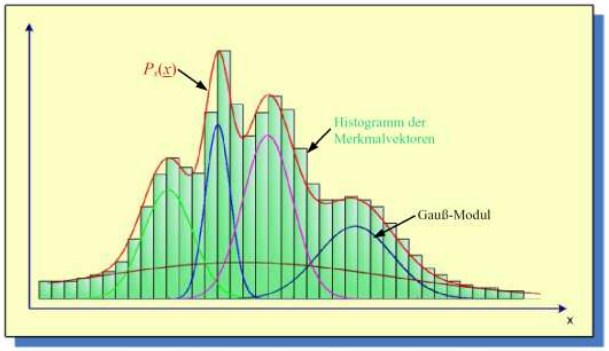
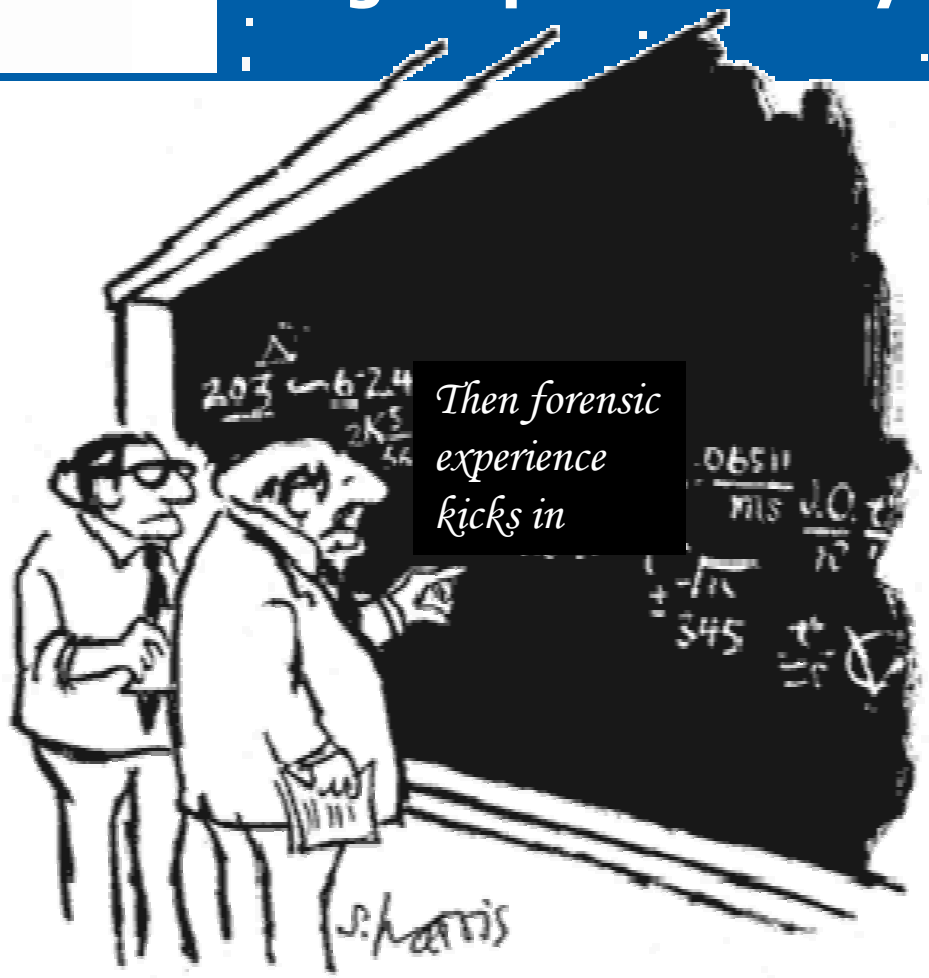


Abbildung 3.5: schematische Darstellung eines GMM-Modells

Possible format of conclusions (classical example; ignoring current discussions on new conclusion schemes)

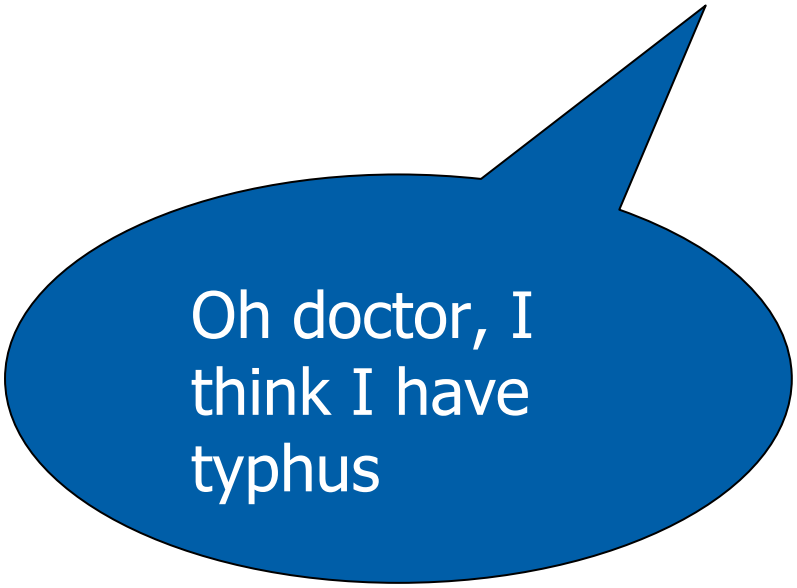
Identity or non-identity
- cannot be given (non liquet)
- applies with predominant probability
- applies with high probability
- applies with very high probability
- applies with a probability close to certainty

Big step from analysis to conclusions



LR approach as a way of making this step more quantifiable and objective

Bayes‘ Theorem: medical example
(mostly hypothetical)



Bayes‘ Theorem: medical example

Posterior odds: Likelihood Ratio: Prior odds:

$$\frac{p(\text{typhus}|\text{red skin})}{p(\neg\text{typhus}|\text{red skin})} = \frac{p(\text{red skin}|\text{typhus})}{p(\text{red skin}|\neg\text{typhus})} \times \frac{p(\text{typhus})}{p(\neg\text{typhus})}$$

Bayes' Theorem: medical example

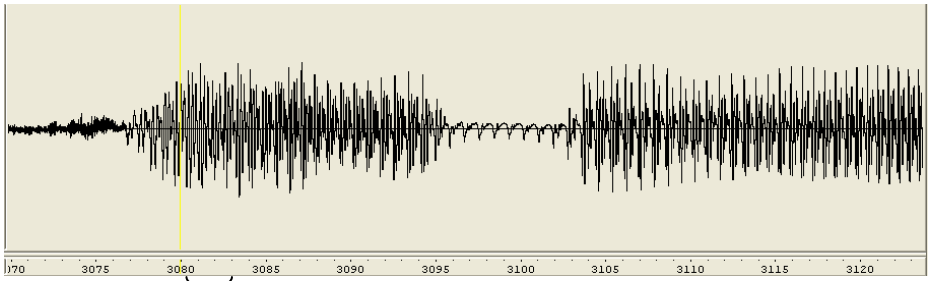
Posterior odds: Likelihood Ratio: Prior odds:

$$\frac{p(\text{typhus}|\text{red skin})}{p(\neg\text{typhus}|\text{red skin})} = \frac{p(\text{red skin}|\text{typhus})}{p(\text{red skin}|\neg\text{typhus})} \times \frac{p(\text{typhus})}{p(\neg\text{typhus})}$$

$$\frac{50}{5000} = \frac{1}{1/50} \qquad \frac{1}{5000}$$ (in Western Europe)

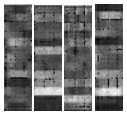
$$\frac{1}{100} \rightarrow$$ Quite low probability that the guy has typhus

Basic principles of an automatic speaker recognition system



Analysis windows

MFCCs



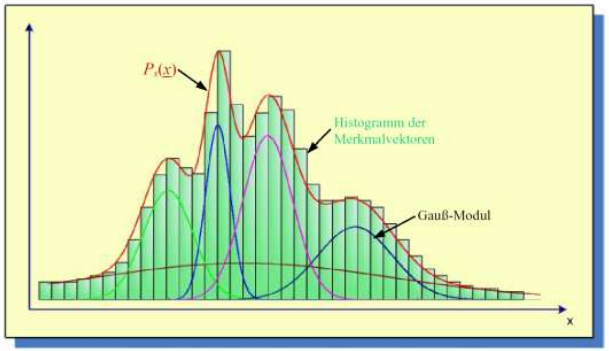


Abbildung 3.5: schematische Darstellung eines GMM-Modells

Step 1: Feature extraction (e.g. MFCC)

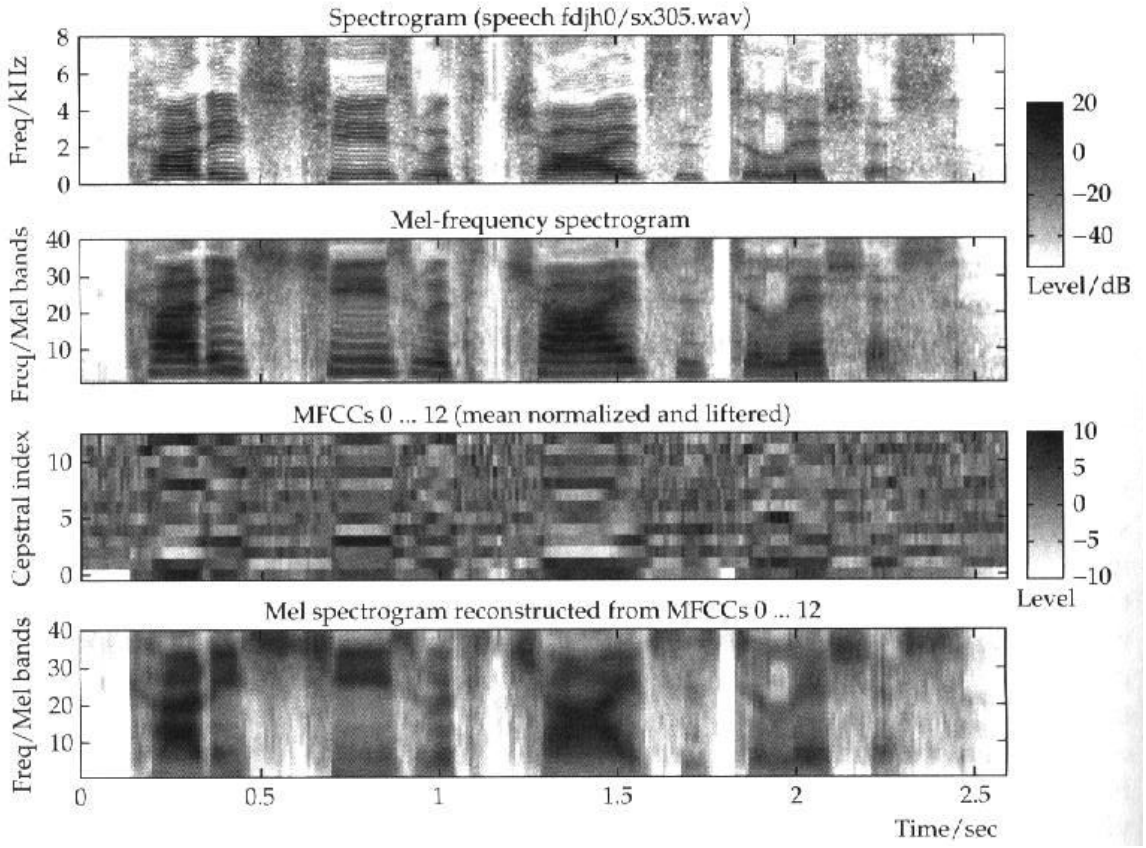
Step 2: Speaker modeling with GMM (suspect and UBM)

Step 3: Calculation of Likelihood Ratios

$$LR = \frac{p(feature\_vectors_{questioned}|GMM_{suspect})}{p(feature\_vectors_{questioned}|GMM_{UBM})}$$

Becker (2012) for more details

MFCC (features used in automatic speaker recognition)

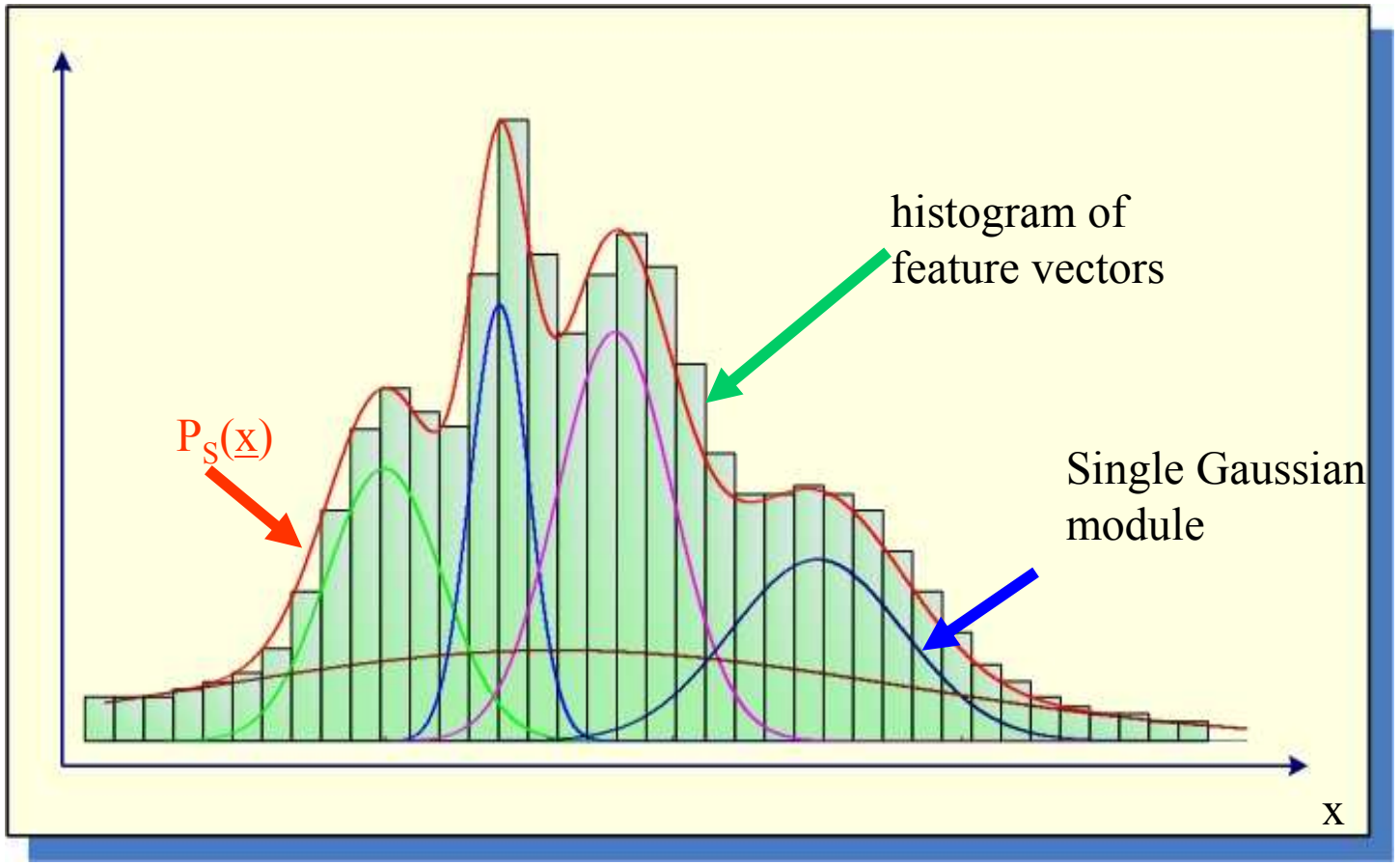


Ellis (2010)

Some characteristics of Mel Frequency Cepstral Coefficients:

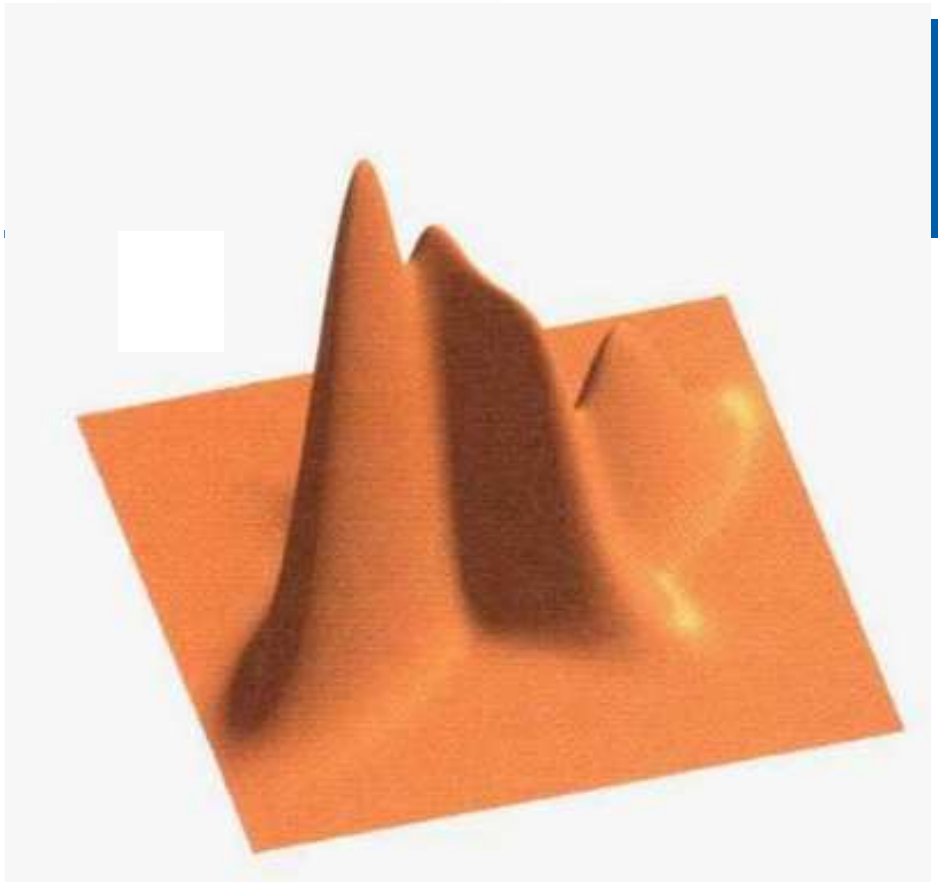
- Focus on the vocal tract properties (filter), ignoring f0 (source)
- The different MFCCs (ca. 13) are (nearly) uncorrelated
- Unsupervised (automatic) extraction: Irrelevant and disturbing acoustic information will be included unless taken care of by further methods (e.g. pause removal, cepstral mean subtraction)

Feature modeling with GMM



For suspect and UBM

Complexity of GMM



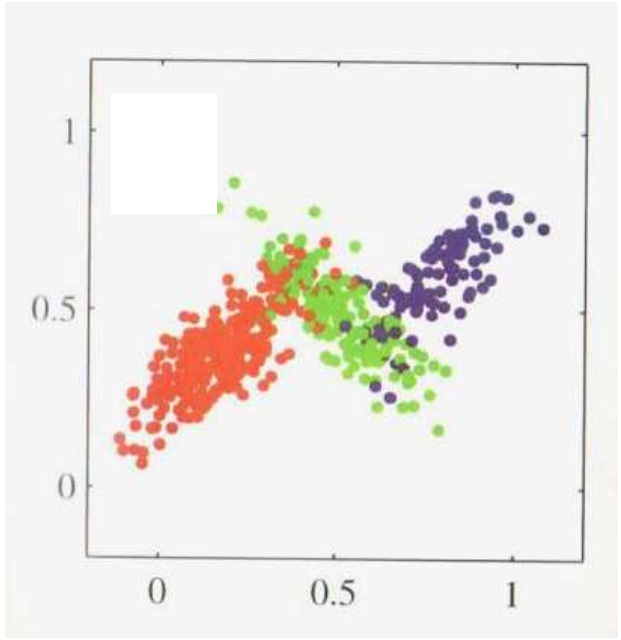


Illustration of 3 Gaussians in a two-dimensional space (Bishop 2006: 112, 433)

Typical automatic system is much more complex:
ca. 32 Gaussians (or more) in a 26-dimensional space (13 static MFCCs, 13 deltas)

Principle of LR calculation I

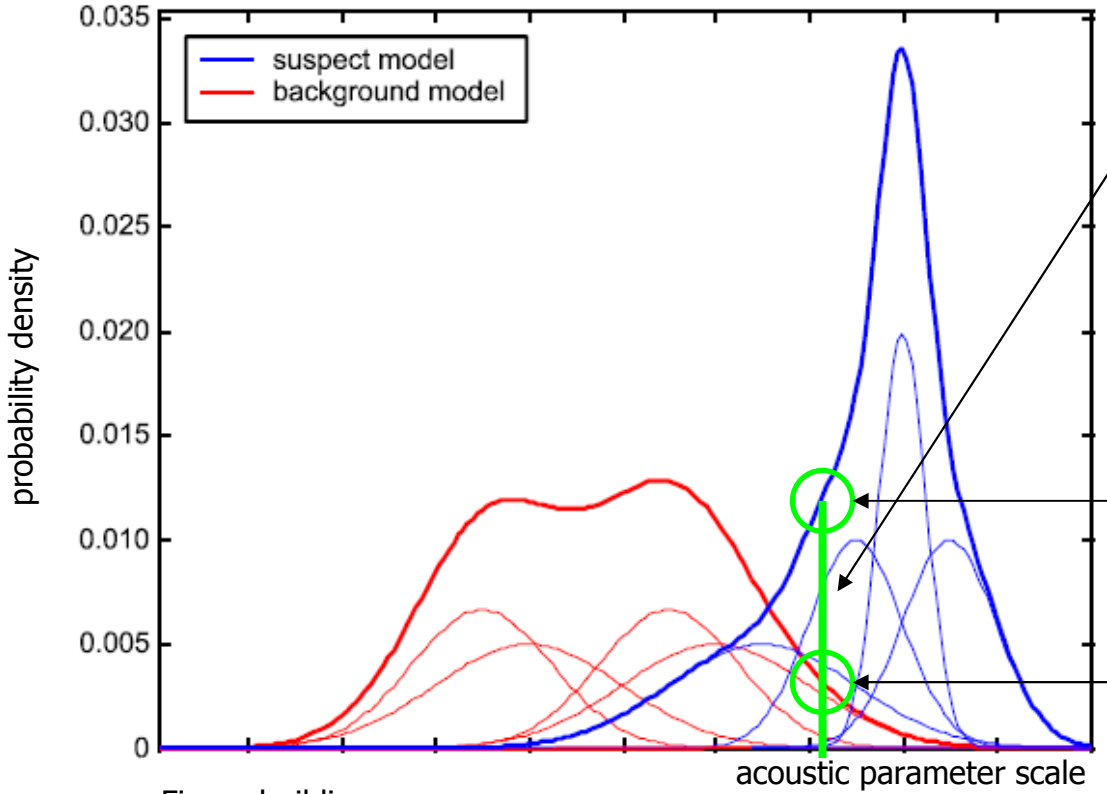


Figure building upon Morrison (2010)

Evidence = feature vector from questioned-speaker recording

$p(feature\_vector_{questioned}|GMM_{suspect})$ (1)

$p(feature\_vector_{questioned}|GMM_{UBM})$ (2)

LR= (1) divided by (2).
here LR>1: support for identity hypothesis

Principle of LR calculation II

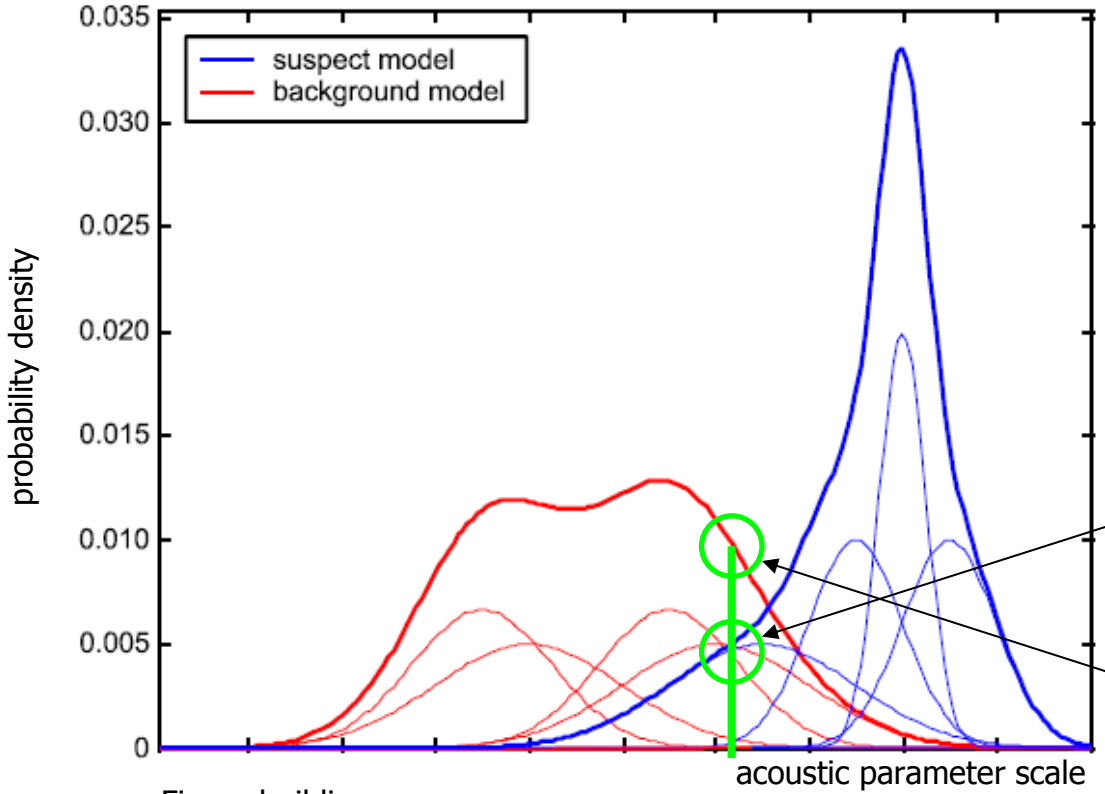


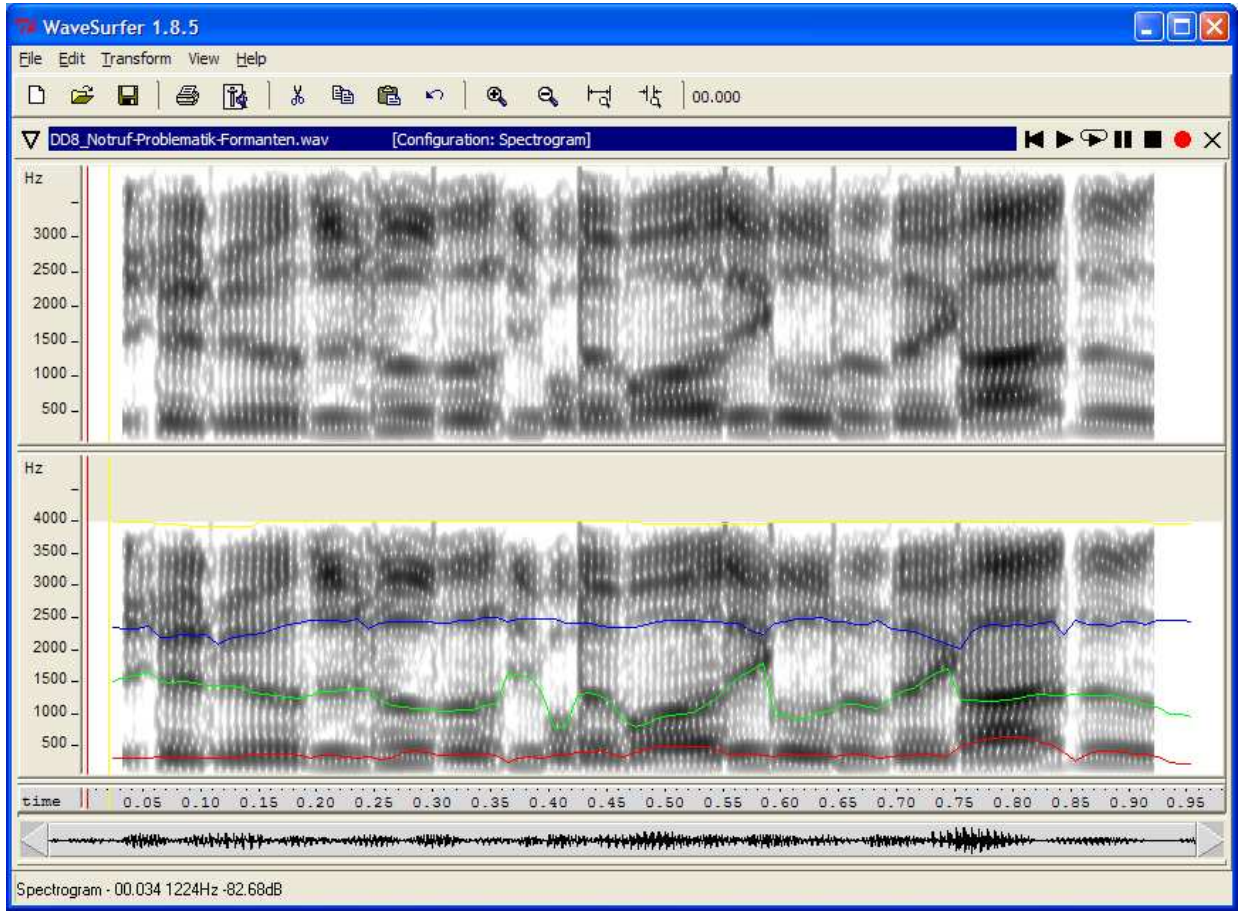
Figure building upon Morrison (2010)

$p(feature\_vector_{questioned}|GMM_{suspect})$ (1)

$p(feature\_vector_{questioned}|GMM_{UBM})$ (2)

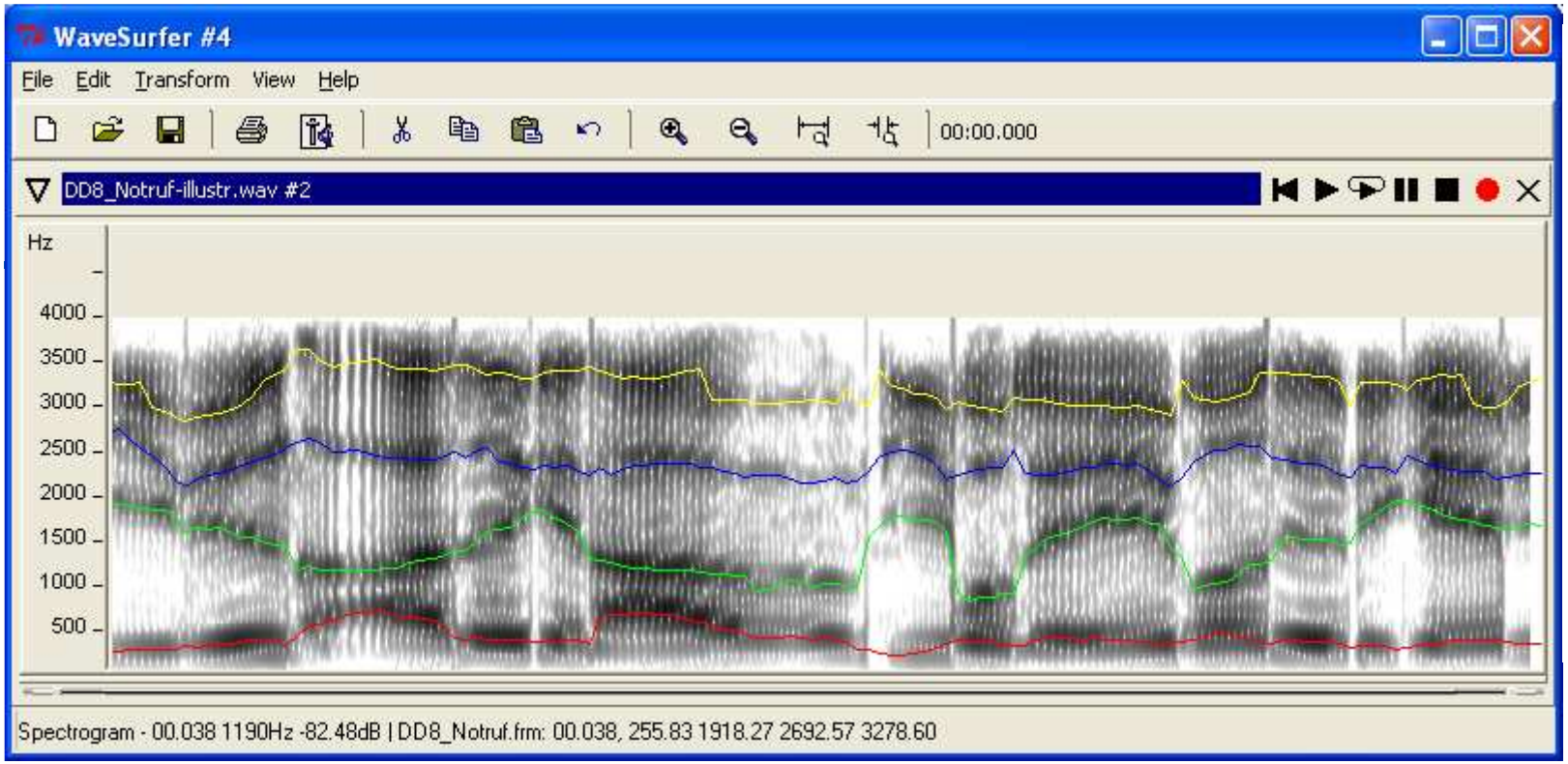
LR= (1) divided by (2).
here LR**<**1: support for **non**-identity hypothesis

Long-Term Formant Distribution (LTF) method (Nolan & Grigoras 2005)



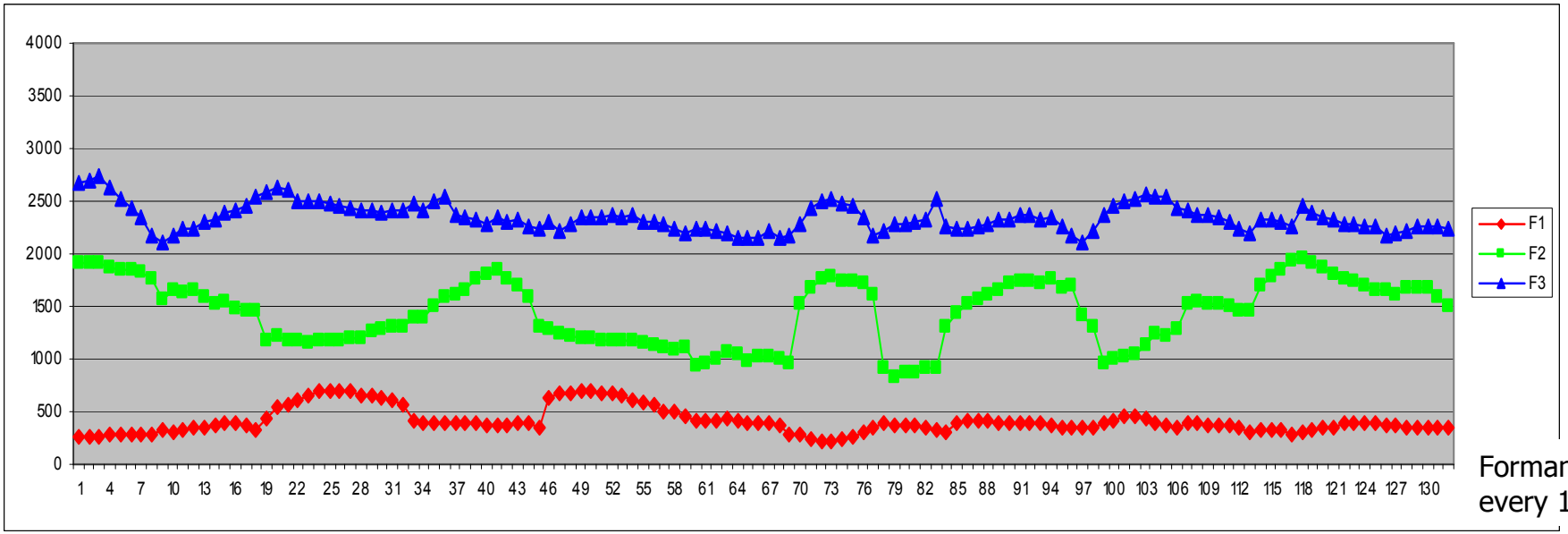
Step 1: Edit the signal in such a way that only vowels with clear formant structure remain

Step 2: LPC-analysis and manual correction of formant tracks



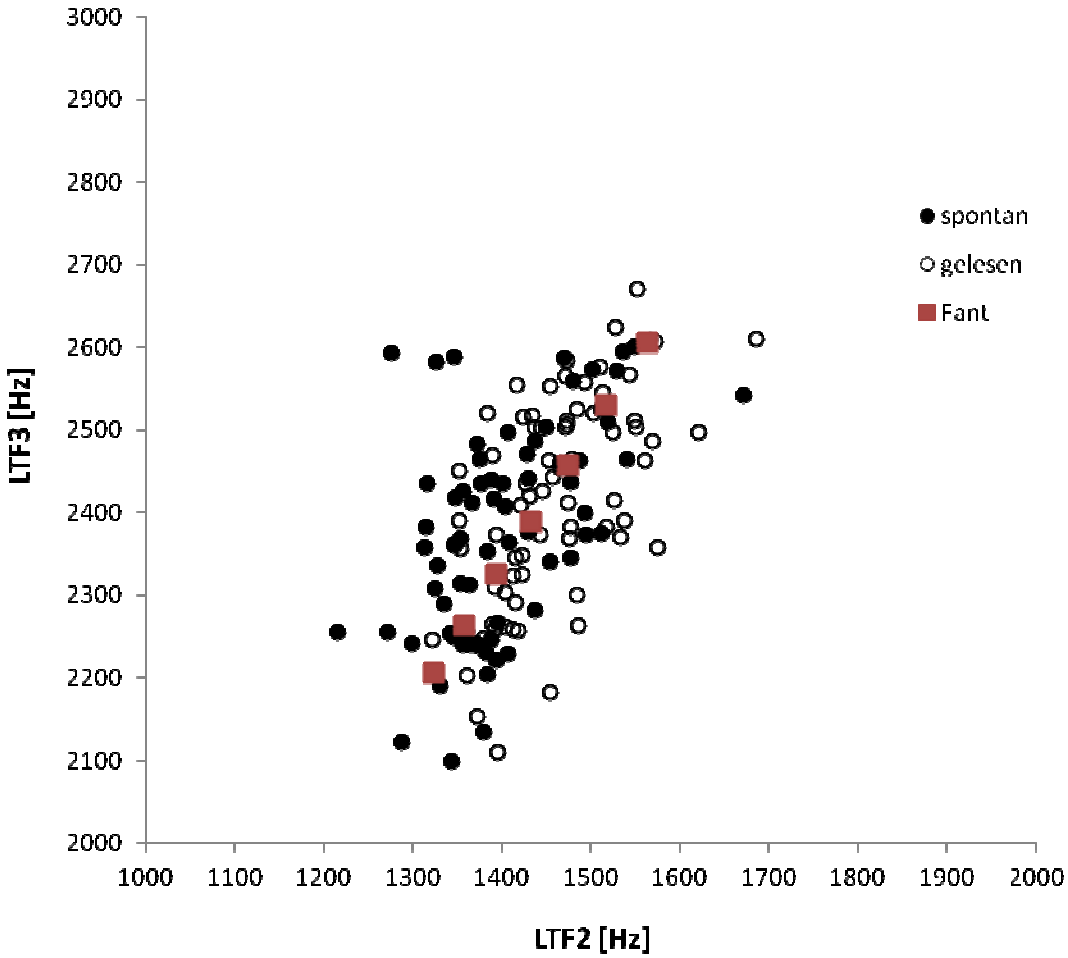
Step 3: Exporting the formant tracks F1,2,3 for further processing

F1 of limited reliability in telephone speech;
F4 unreliable or invisible



Formant values every 10 ms

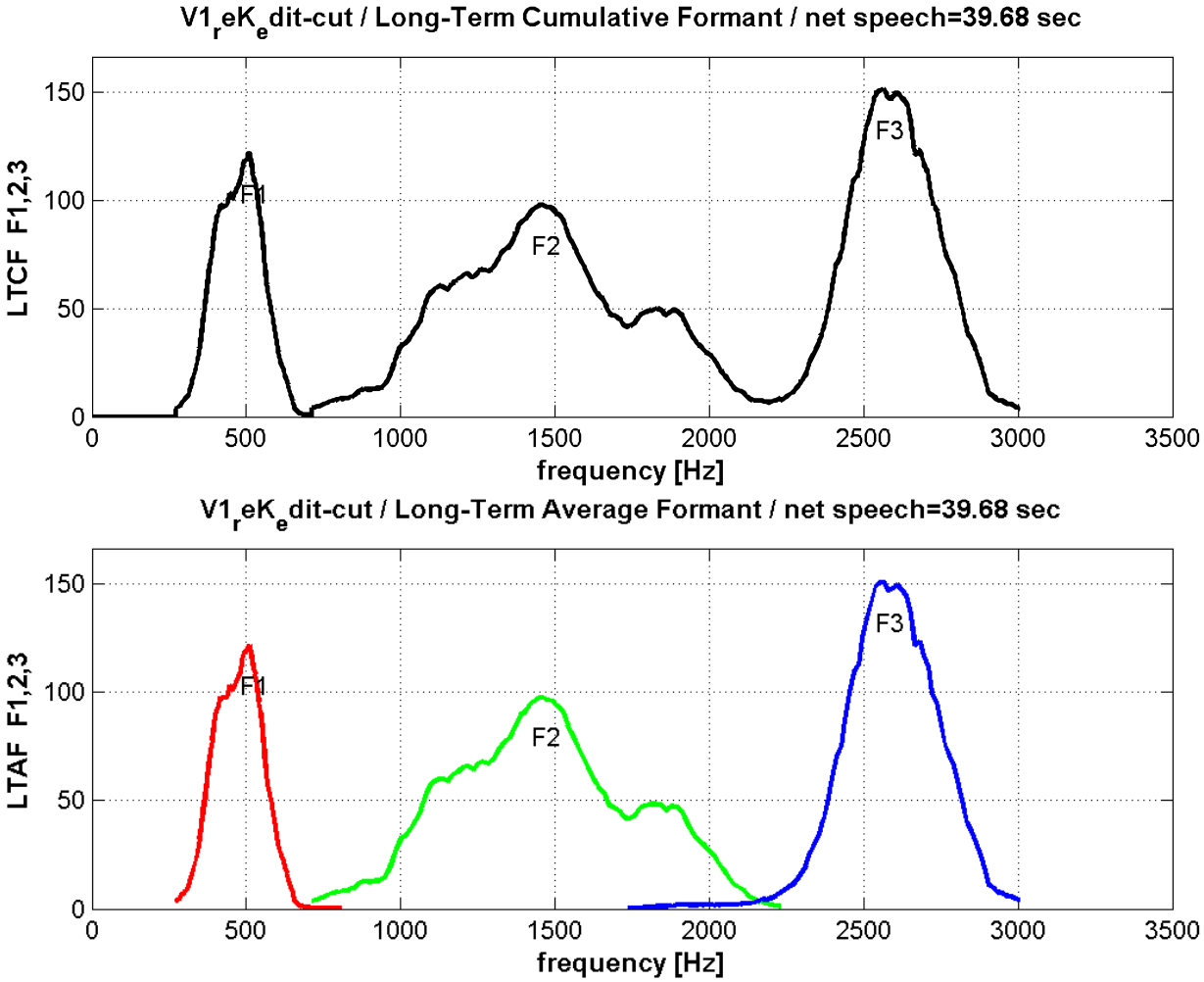
LTF reference data on German



Measurement results from 71 male adult speakers of German (Moos 2008)

and correspondence with predictions by Fant (1960), based on the well-known tube model

LTF-based histogram



Printout from software created by Catalin Grigoras

LTF-data yield complex distributions.

GMM modeling seems like a good idea.

First work in this area: Becker et al. (2008)

Comparison of methods (example; exact values can differ depending on the system)

	Automatic speaker recognition	Long-term formant analysis
Features	MFCC	Formants (frequencies and possibly bandwidths)
Number of feature dimensions	13 (without Deltas)	3 (F1, F2, F3) (plus possibly formant bandwidths)
Number of Gaussians	32	8
Delta coefficients	established	open to research
Channel normalisation	established	open to research
Performance on good-quality data	more	less
Theory-drivenness; „Anschaulichkeit" (Phil Rose, p.c.)	less	more
Robustness against mismatch	limited	open to research

Conclusion

- LR approach flexible enough to encompass both automatic (speech technological) and phonetic (esp. acoustic phonetic) methods
- Within LR approach, several further methodological parallelisms in the details (e.g. GMM modeling; delta coefficients). Interesting perspectives for research and practice.
- Further research:
  - Combining the results from automatic and phonetic methods ("fusion")
  - Applying the methods to forensically authentic data
  - Speaker-discrimination performance of automatic and phonetic methods under mismatched conditions

References

Becker, Timo (2012): *Automatischer forensischer Stimmenvergleich*. Norderstedt: Books on Demand.

Becker, Timo, Michael Jessen & Catalin Grigoras (2008): Forensic speaker verification using formant features and Gaussian mixture models. *Proceedings of INTERSPEECH 2008*, Brisbane, pp. 1505-1508.

Bishop, Christopher M. (2006): *Pattern Recognition and Machine Learning*. Berlin: Springer.

Ellis, Daniel B. (2010: An introduction to signal processing for speech,. In: William J. Hardcastle, John Laver and Fiona Gibbon (eds.) *The Handbook of Phonetic Sciences, second edition*. Chichester: Wiley-Blackwell. 757-780.

Fant, Gunnar (1960): *Acoustic Theory of Speech Production*. The Hague: Mouton.

Jessen, Michael (2012): *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*. München: LINCOM.

Moos, Anja (2008): Forensische Sprechererkennung mit der Messmethode LTF (long-term formant distribution). MA thesis, Universität des Saarlandes. www.psy.gla.ac.uk/docs/download.php?type=PUBLS&id=1286.

Morrison, Geoffrey Stewart (2010): *Forensic Voice Comparison*. In: I. Freckelton & H. Selby (eds.) Expert Evidence (Chapter 99). Sydney: Thomson Reuters.

Nolan, Francis and Catalin Grigoras (2005): A Case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law* 12: 143-173.