Bundeskriminalamt

Long-Term Formant Distribution as a forensic-phonetic feature

ASA 2nd Pan-American/Iberian Meeting on Acoustics

Cancún, México, Nov 15-19, 2010

Michael Jessen and Timo Becker
BKA, Department of Speaker Identification and Audio Analysis (KT54)

Structure

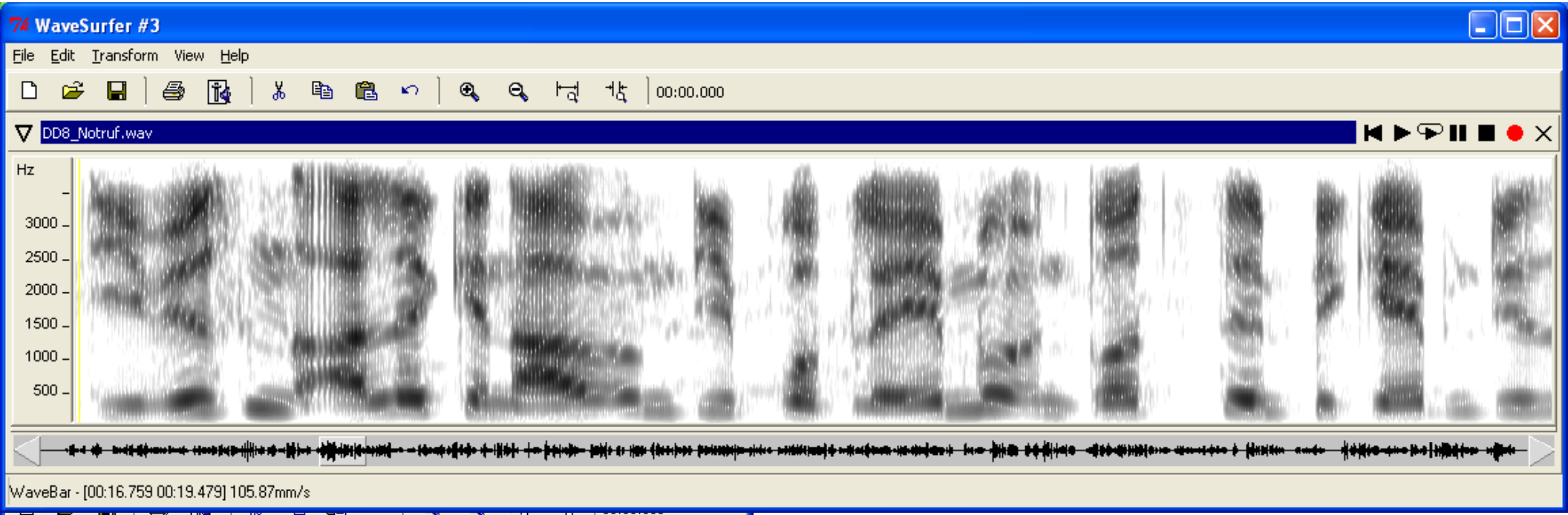
1. Long-Term Formant Distribution: measurement methods and background
2. LTF and body height
3. LTF measurement consistency
4. Language dependence of LTF
5. Recognition performance based on LTF and automatic speaker recognition
6. Conclusions

Long-Term Formant (LTF) Distribution: terminology

Long-Term Formant Distribution (Nolan & Grigoras, 2005) is a global (as opposed to segment-based) representation of vowel formant frequencies over an entire recording of a speaker (or over a long stretch of speech from that speaker).

Formant frequencies are extracted with a formant tracker (LPC-based) and manually corrected. No segmentation into sounds is performed.

The resulting distribution of formant values (mainly F2 and F3) can be characterized in different ways. The simplest way is to calculate the average. More advanced ways include modeling of the LTF distribution with Gaussian Mixture Models (GMM) (Becker et al., 2008).



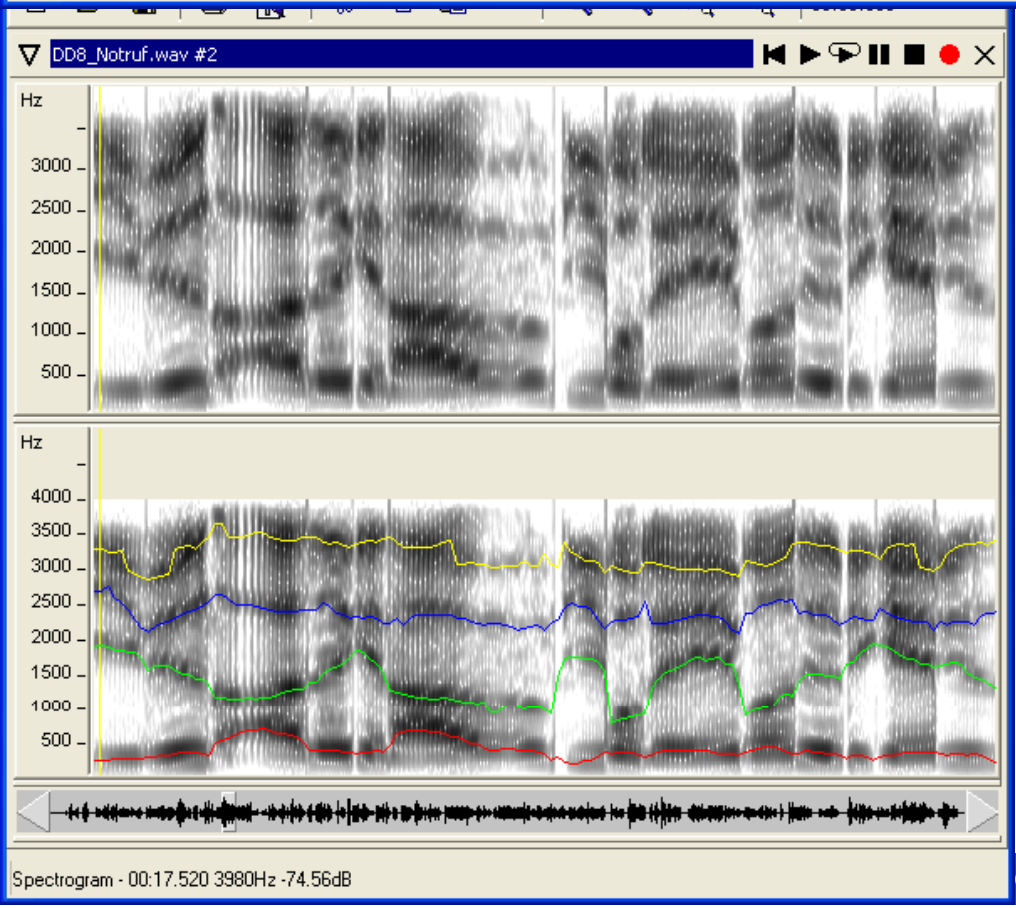
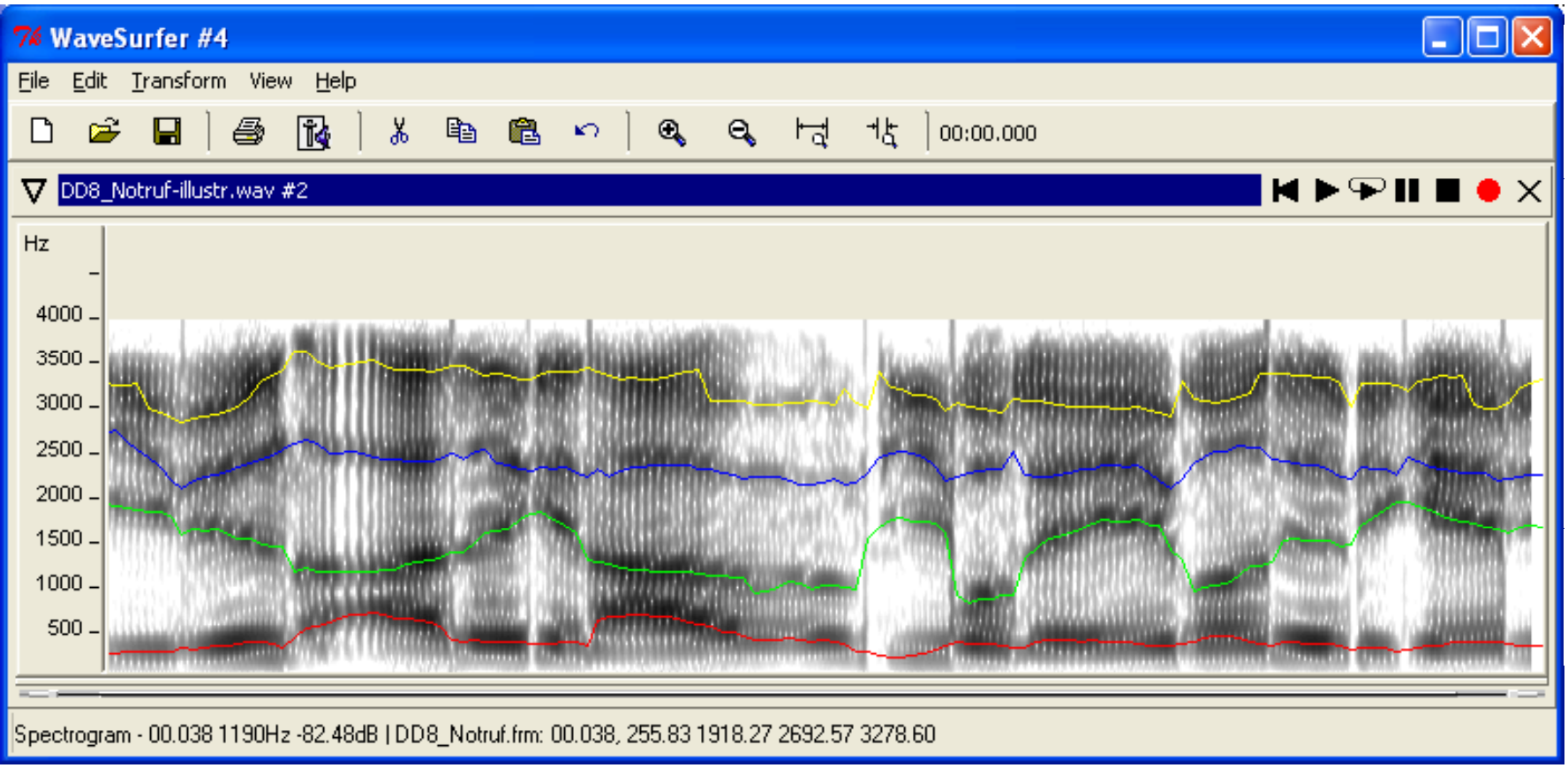


Illustration of the method:

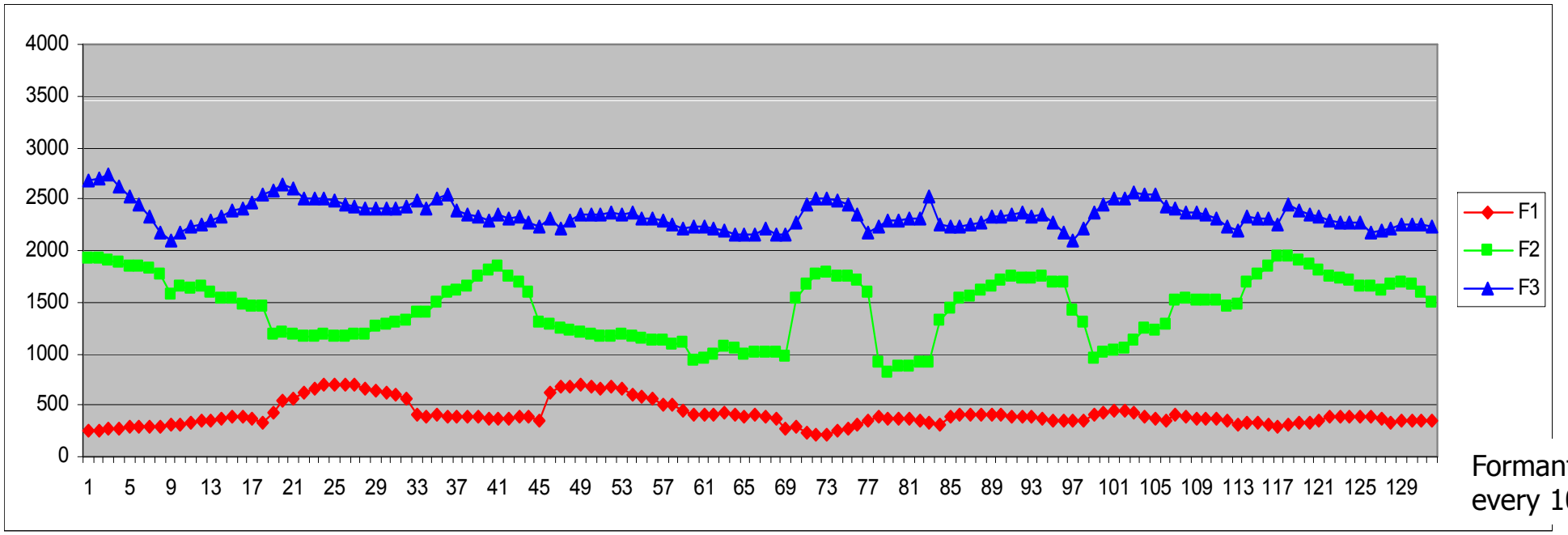
Step 1: Editing the signal in a way that only vowels with clear formant structure remain

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



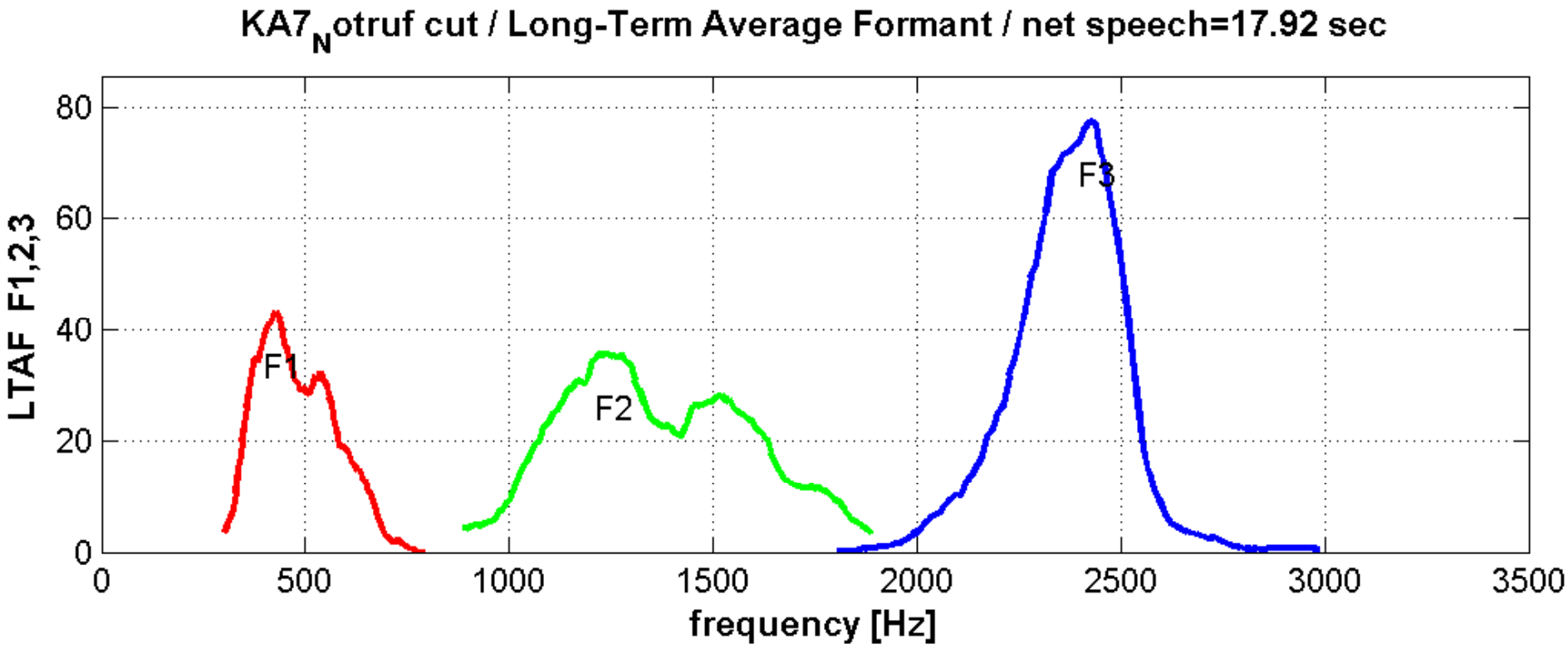
Step 3: Exporting the formant tracks F1,2,3 for further processing

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



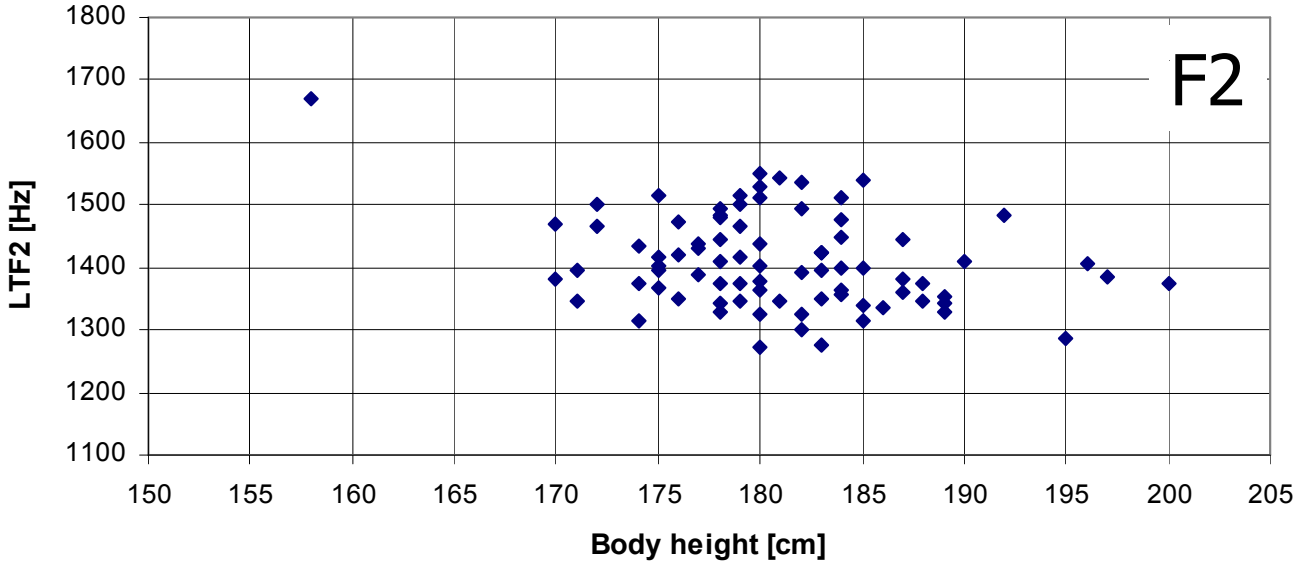
Formant values every 10 ms

Example of the raw LTF distribution of a speaker



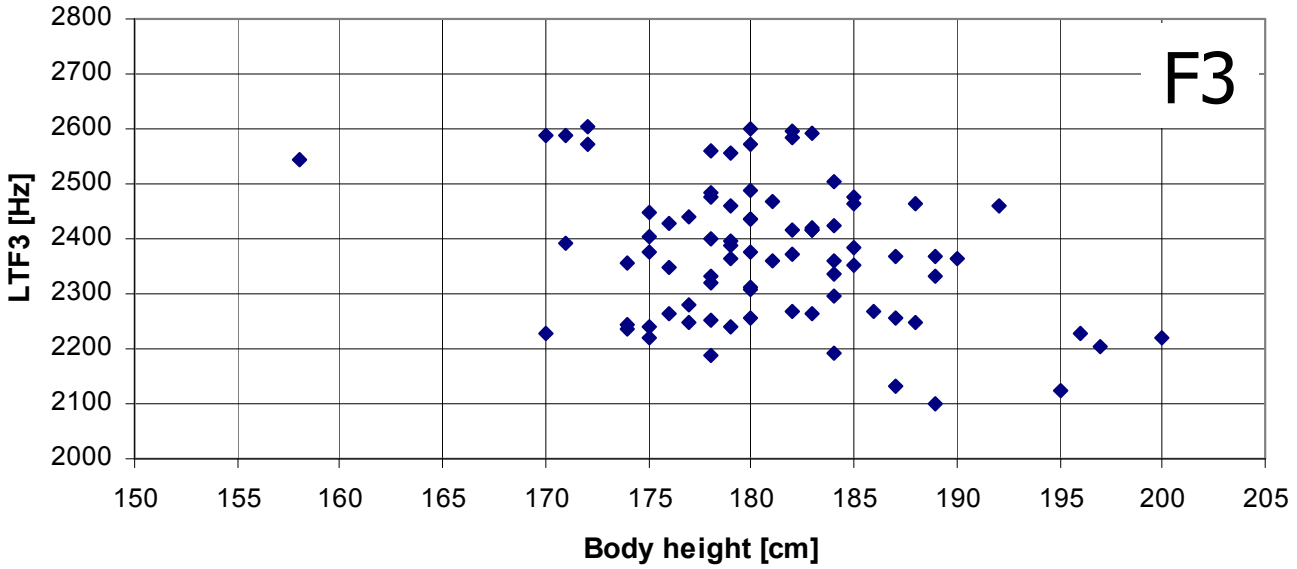
from freeware *Catalina Forensic Expert opinion v1.0*
from Catalin Grigoras (U Colorado Denver)
http://www.forensicav.ro/download/CatalinaManual3h.pdf

Correlation between LTF and body height



Pearson's product-moment correlation

One-sided (less)
rho=-0.315726857072528
p=0.00204454743894922

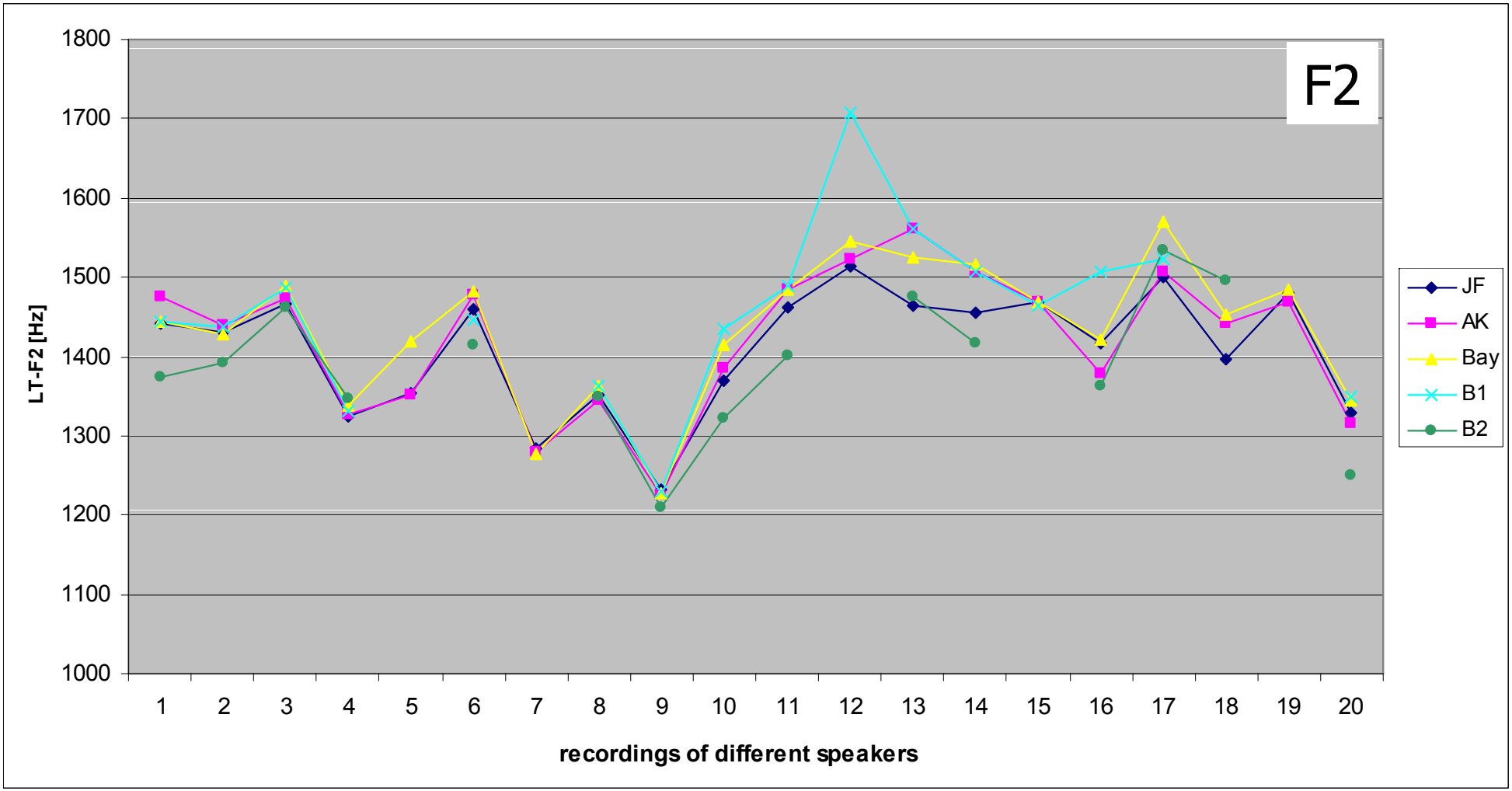


rho=-0.339139631480740
p=0.00097693931875183

Significant negative correlations between long-term formant frequencies (F2, F3) and body height

LTF-means from 81 speakers in Pool 2010 (telephone-transmitted) (thanks to Hanna Feiser for assistance)

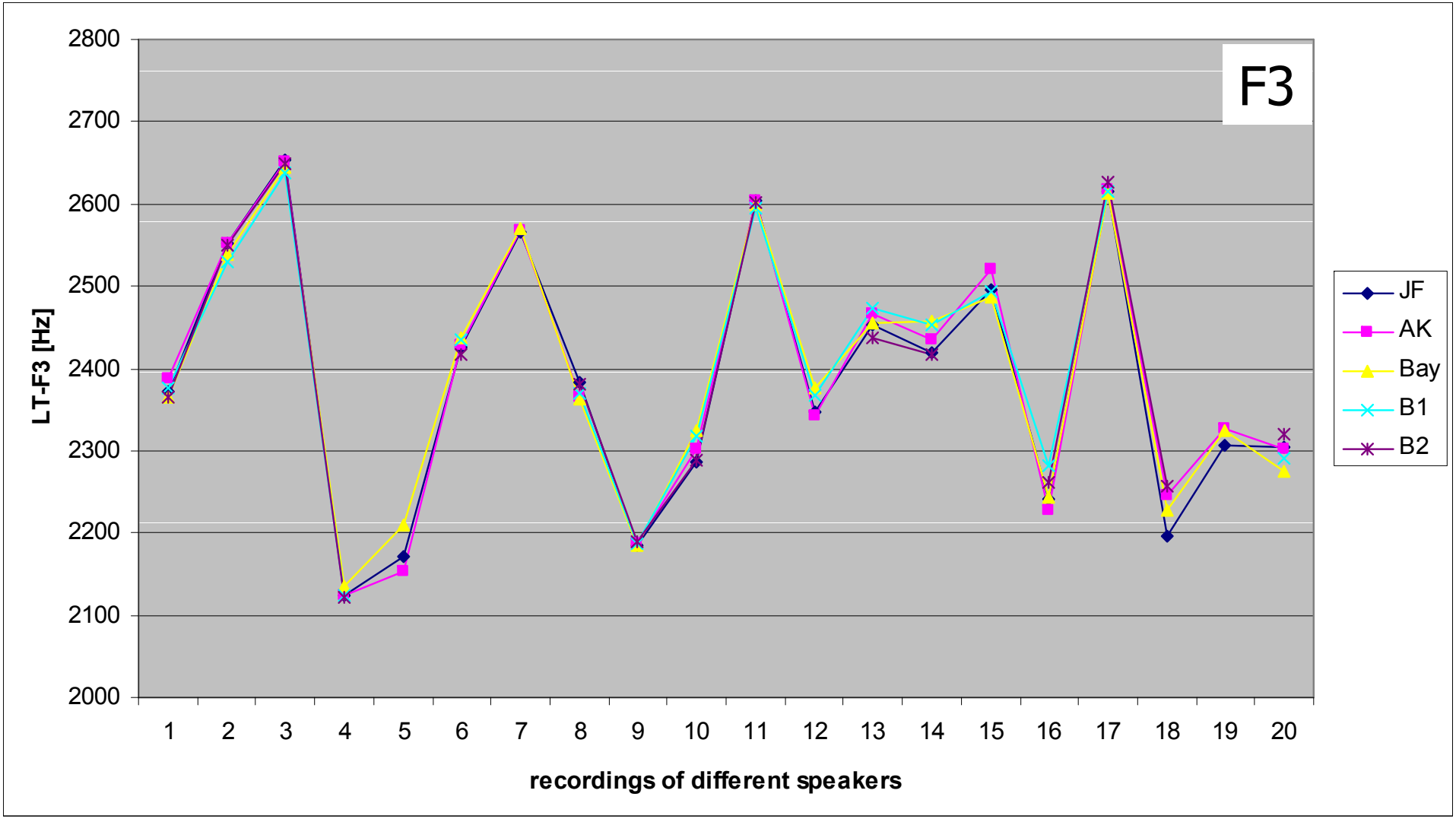
Measurements consistency across phoneticians: LT-F2



Pearson correlations (two-sided) between 0.84 and 0.95

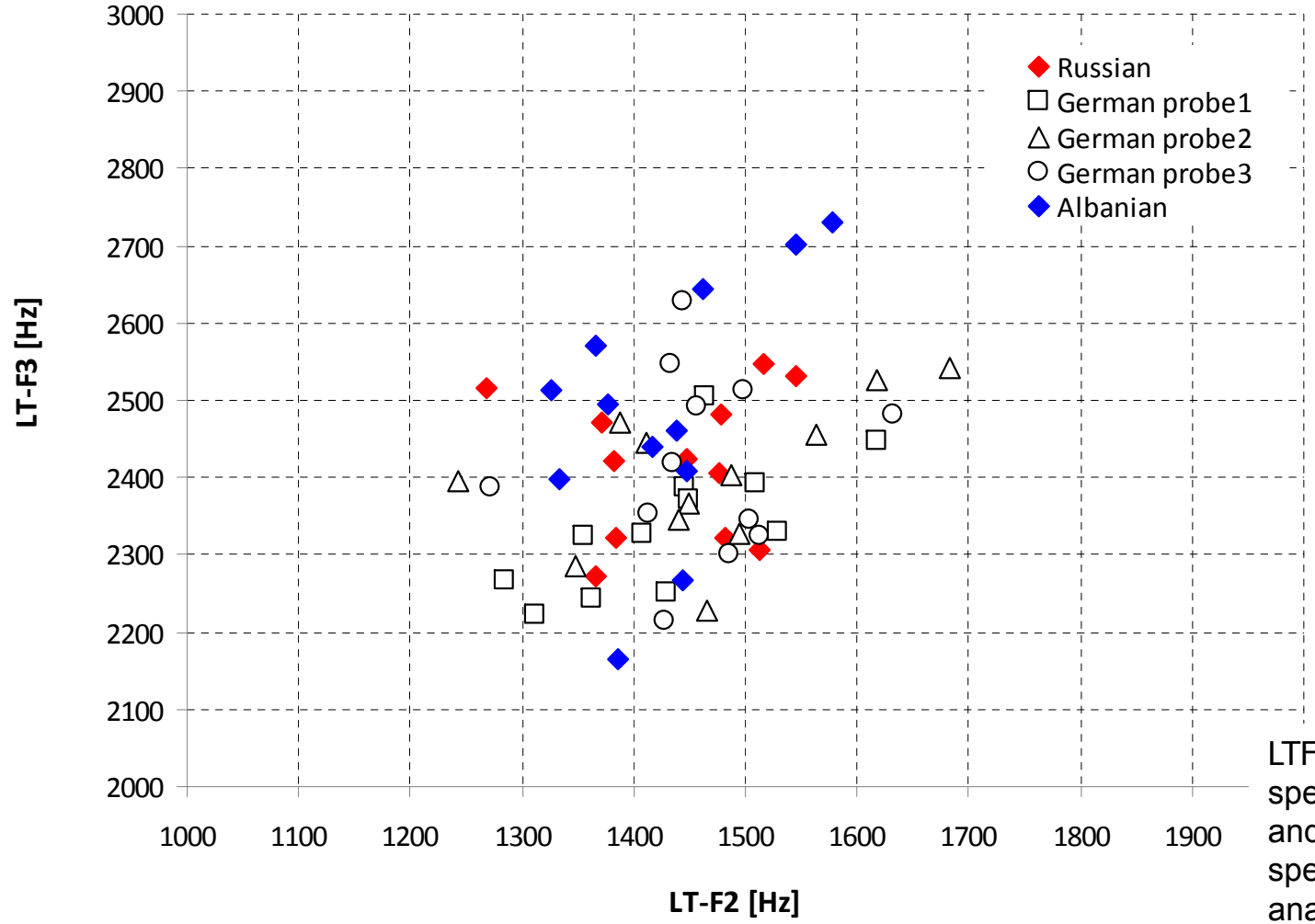
LTF-means from 20 speakers in “Digs” dialect corpus under forensically realistic conditions

Measurements consistency across phoneticians: LT-F3



Pearson correlations (two-sided) between 0.98 and 0.99

Language influence on LTF



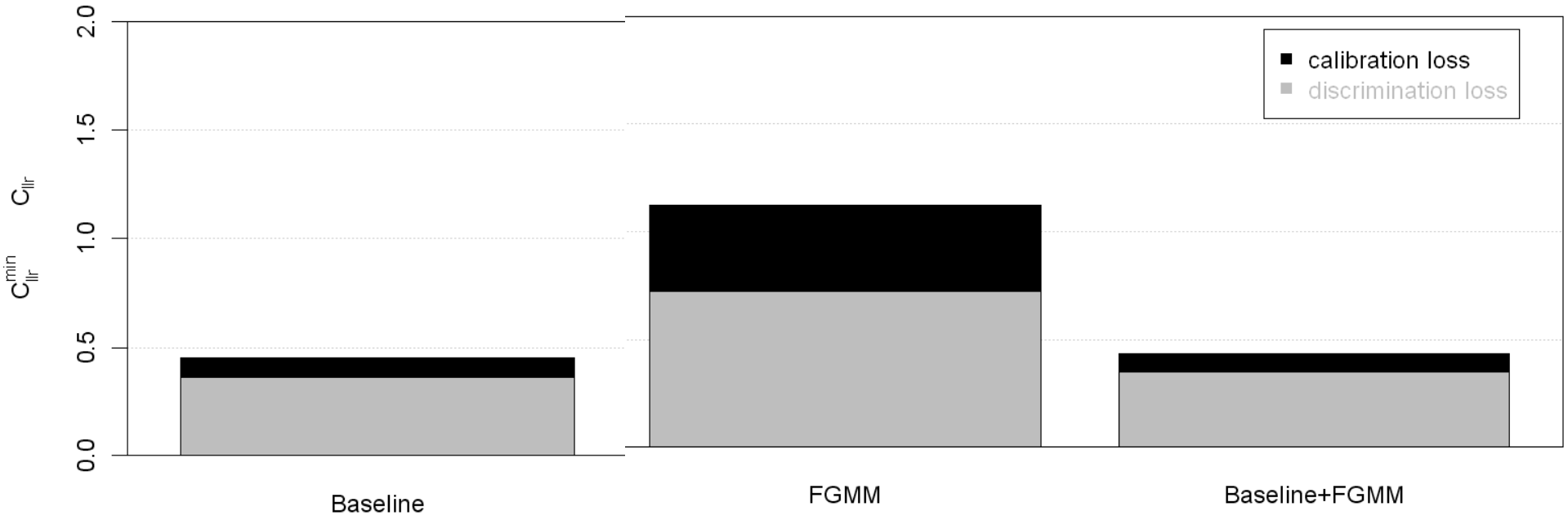
For these data, different languages do not differ in the LTF-space that they occupy (one-way ANOVA [$F(4,55) = 0.44$; $p = 0.77$]).

LTF-means from three German speakers in Digs dialect corpus and from Russian and Albanian speakers in case data under analogous conditions (spont telephone)

Speaker recognition tests

37 target trials and 803 non-target trials, involving 21 speakers from casework, comparing:

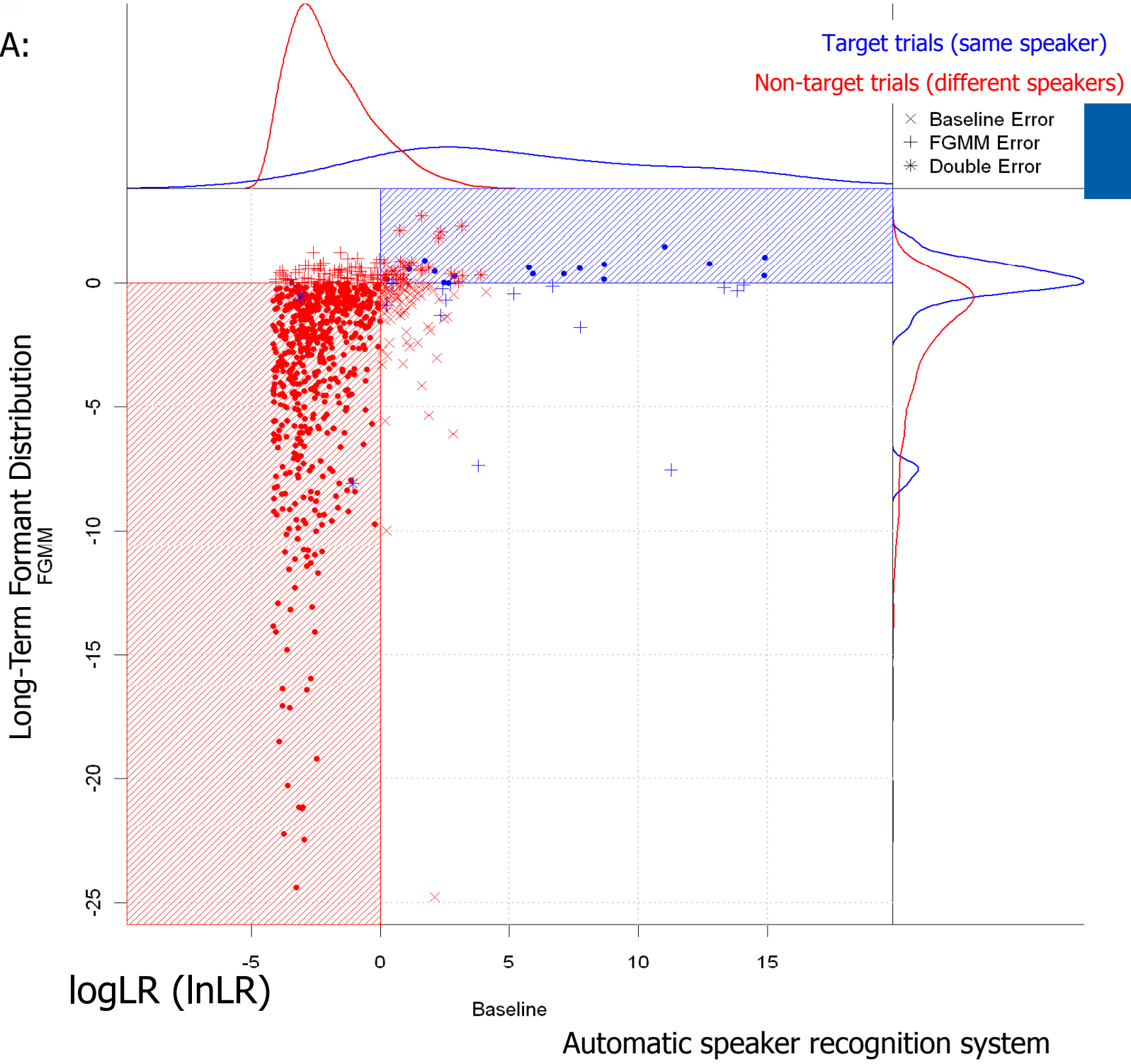
- *Baseline* = a standard GMM-UBM automatic system
- FGMM = GMM-modeled LTF



New development at BKA:

DiSC-Plot

Discrimination, Scatter, Correlation



Conclusions: LTF analysis in forensic phonetics and acoustics (1)

- ☺ LTF (F2 and F3) correlates negatively with body height (relevant for voice profiling).
- ☺ LTF measurements have high consistency across phonetic experts.
- ☺ Pending further tests and with some degree of caution, LTF statistics established for one language can be used across languages.
- ☺ LTF (F2 and F3) do not differ much between different vocal effort levels. Vocal effort differences are a common problem in forensic material.

Conclusions: LTF analysis in forensic phonetics and acoustics (2)

☹ Performance of LTF analysis with classical evaluation measures (DET-plots, APE-plots, C_{llr}) is worse than performance of automatic speaker recognition and fusion does not increase overall performance. But:

😐 The tests so far are based predominantly on matching conditions; under mismatched conditions, the relative performance of LTF analysis might increase.

☺ Detailed results in the DiSC plot shows that LTF and automatic speaker recognition can make different errors: using both methods is a good safeguard against false conclusions.

☹ Quite limited LR values in same-speaker comparisons (max about LR=16 in case material for the tests so far): LTF cannot give very strong support for same-speaker hypothesis.

☺ Different-speaker comparisons can yield very low LR values: LTF can give very strong support for different-speaker hypothesis.

References

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

Kirchhübel, Christin (2009): The effects of Lombard speech on vowel formant measurements. MSc thesis, University of York, UK.

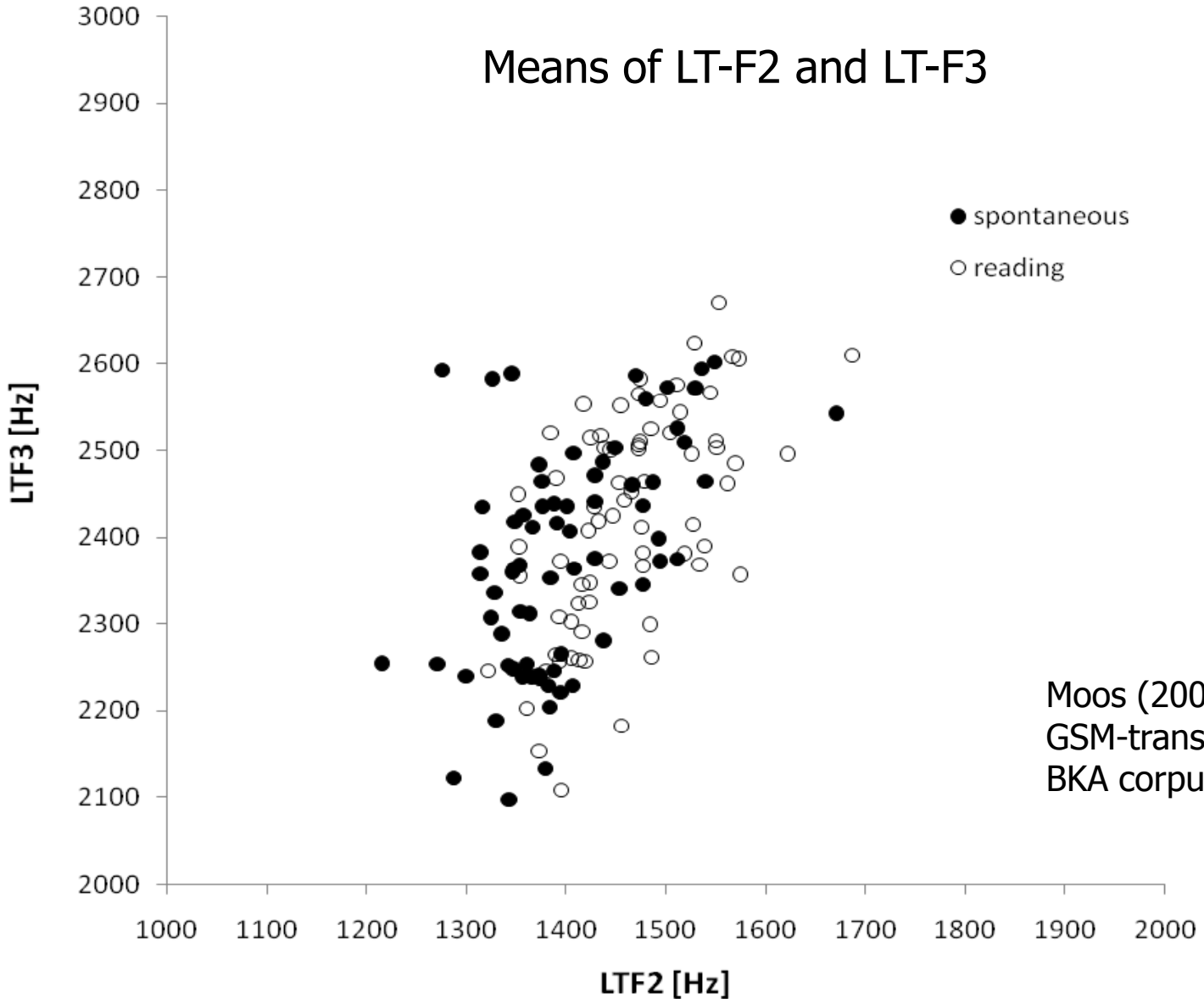
Moos, Anja (2008): Forensische Sprechererkennung mit der Messmethode LTF (long-term formant distribution). MA thesis, Universität des Saarlandes. www.psy.gla.ac.uk/docs/download.php?type=PUBLS&id=1286.

Moos, Anja (2010): Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. To appear in *The Phonetician*.

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.

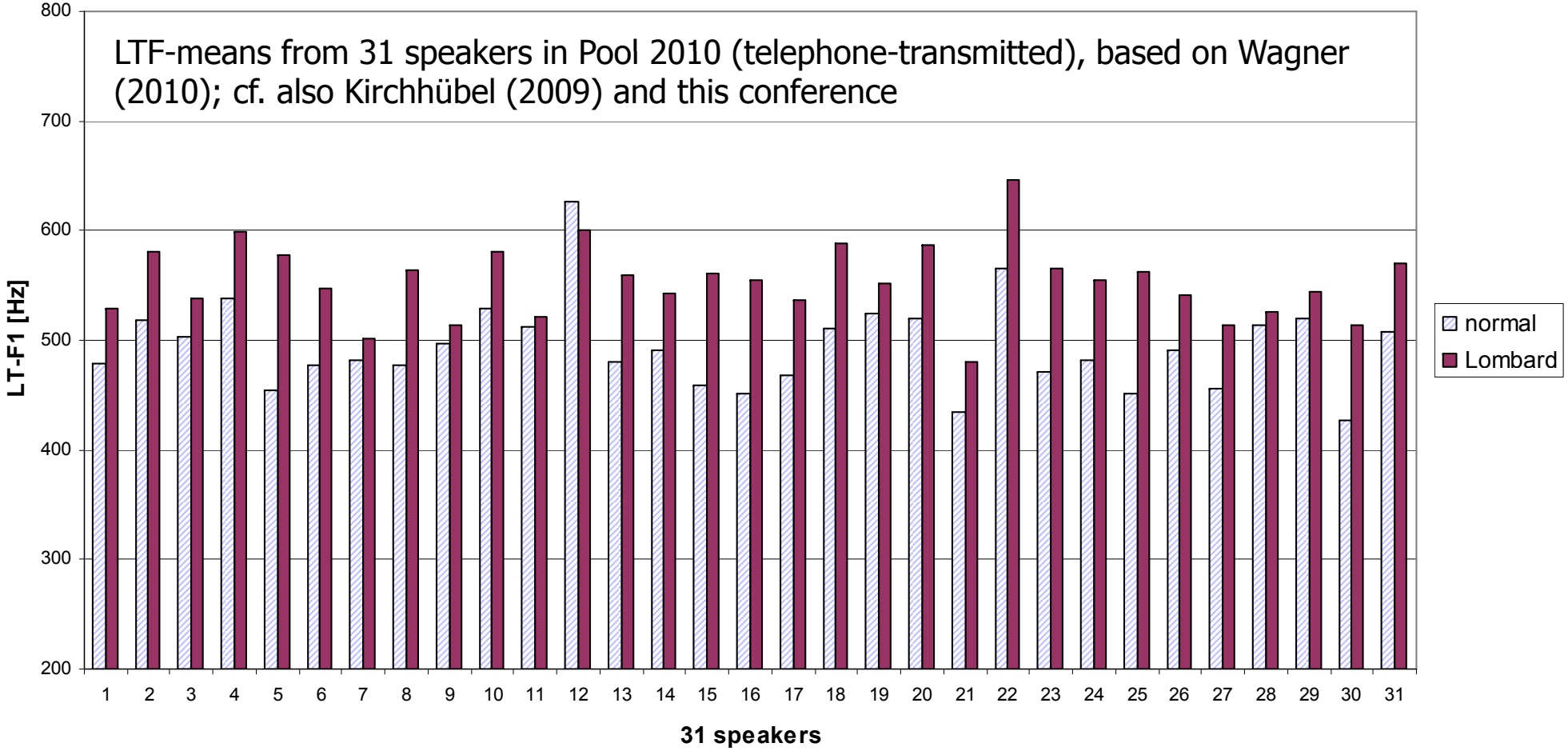
Wagner, Katrin (2010): Der Einfluss der Sprechlautstärke auf die ersten drei Vokalformanten in mobilfunkübertragener Sprache: Forensischer Stimmenvergleich anhand der LTF-Methode". BA thesis, Universität Frankfurt.

Inter-speaker variation: Mean LTF for 71 adult male speakers of German



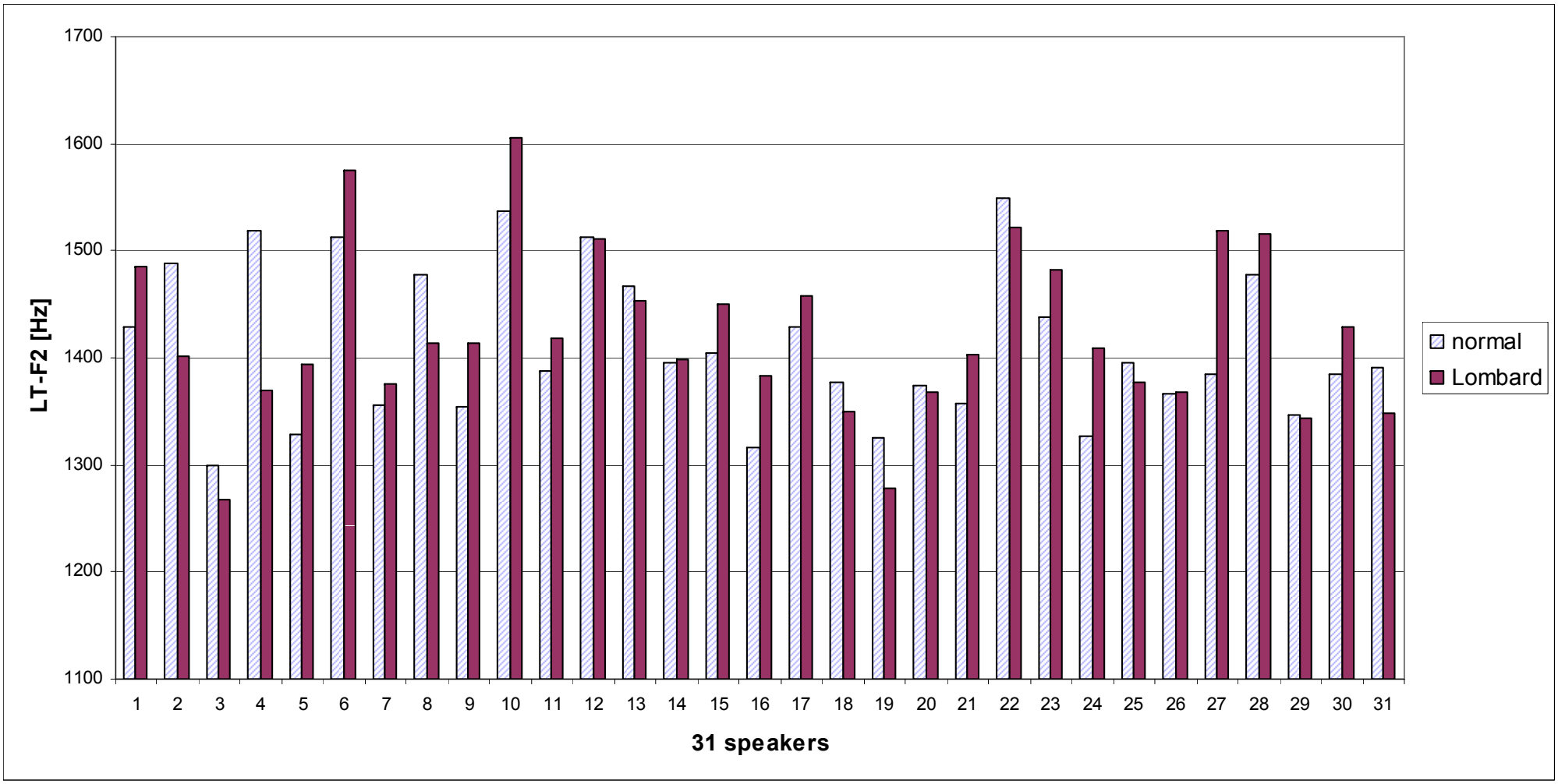
Moos (2008, 2010), based on GSM-transmitted speech in BKA corpus "Pool 2010"

Influence of vocal effort (Lombard condition) on LT-F1



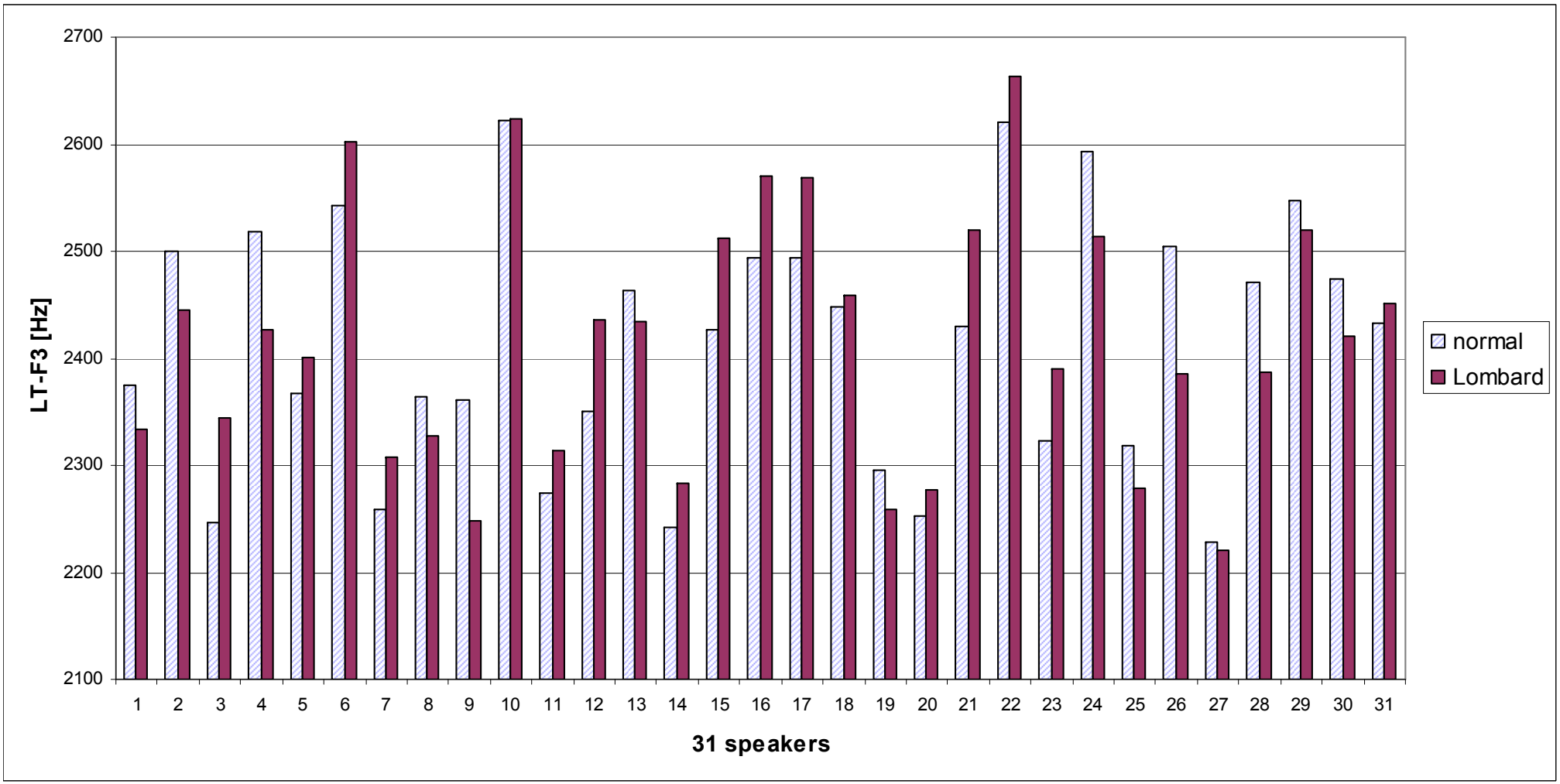
LT-F1 consistently higher in Lombard speech. Significant difference with paired t-test, indicating substantial intra-speaker variation. But: LT-F1 is of limited forensic use anyway (due to the effect of telephone transmission on F1)

Influence of vocal effort (Lombard condition) on LT-F2



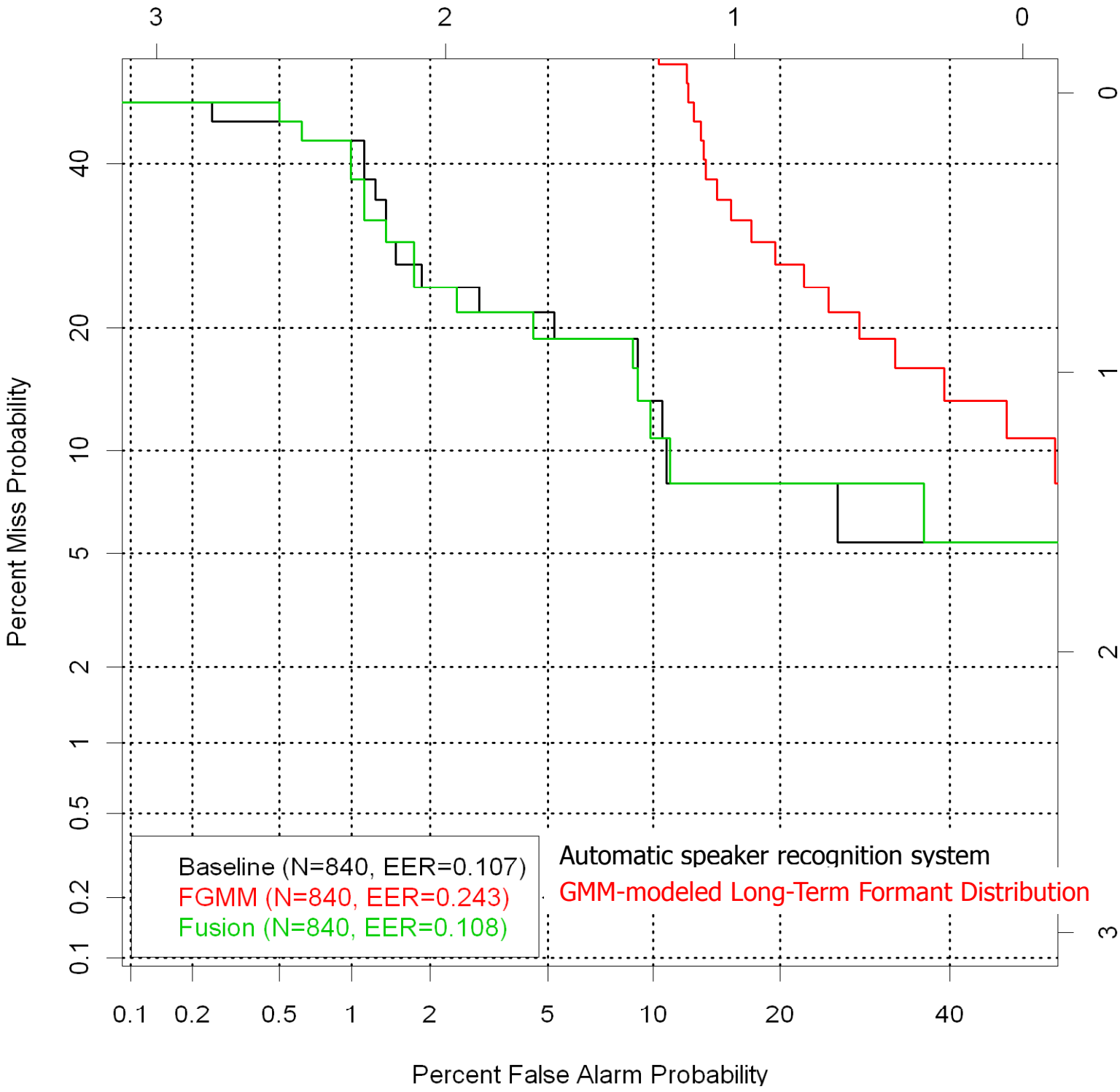
Lombard-effect on LT-F2 inconsistent across speakers. Non-significant difference with paired t-test, indicating acceptable intra-speaker variation.

Influence of vocal effort (Lombard condition) on LT-F3

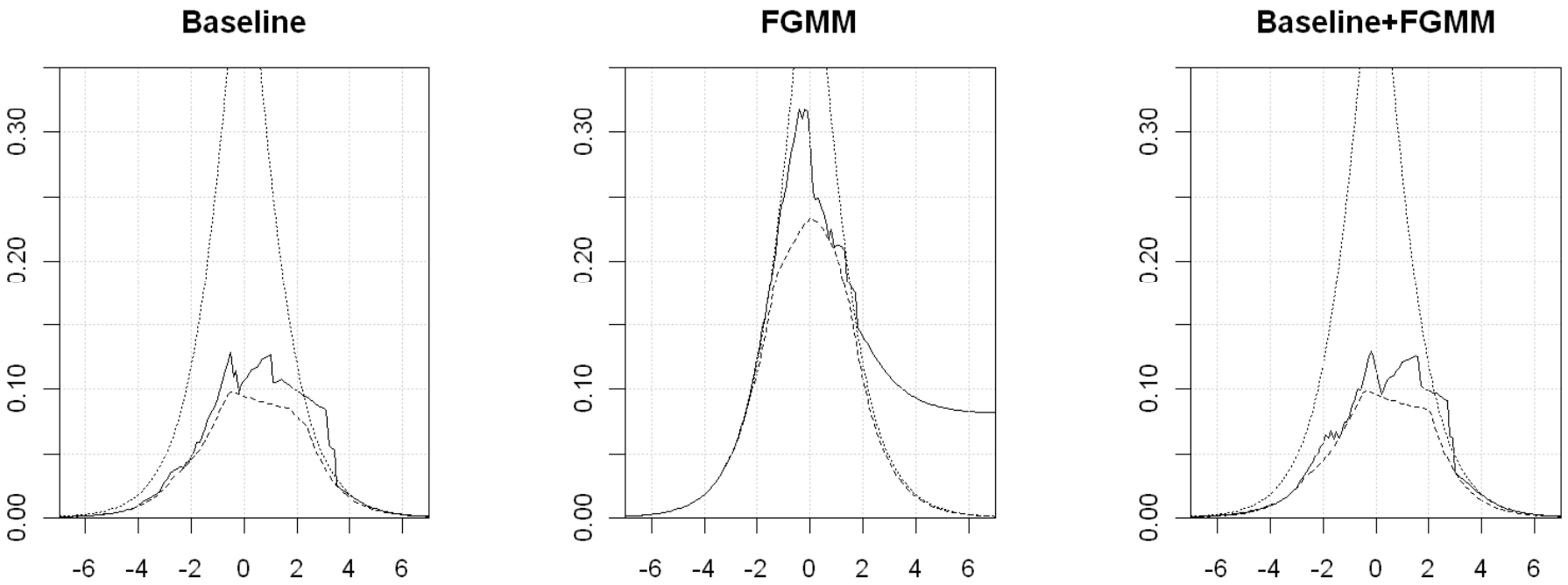


Like with LT-F2: Lombard-effect on LT-F3 inconsistent across speakers. Non-significant difference with paired t-test, indicating acceptable intra-speaker variation.

DET-Plot


Percent Miss Probability
Percent False Alarm Probability
0.1
0.2
0.5
1
2
5
10
20
40
Baseline (N=840, EER=0.107)
FGMM (N=840, EER=0.243)
Fusion (N=840, EER=0.108)
Automatic speaker recognition system
GMM-modeled Long-Term Formant Distribution

APE-Plot



C_{llr}