Wiktor Gonet

APAP

21-23 JUNE 2013

PHONETIC AND PHONOLOGICAL PARAMETERS IN INDIVIDUAL SPEECH CHARACTERIZATION

UMCS

HOW DOES BRAIN WORK IN SPEAKER CHARACTERIZATION?

Speaker characterization is a very complex process and research has barely begun to comprehend what lies behind the easiness of recognizing speech ad voices of the people we know, or what underlies remembering the voices and speech manner of other people so that we can, with time, learn to recognize their voices among those of other people.

Certainly, the speech signal contains two kinds of information: linguistic, allowing us to understand what is being spoken, and speaker-oriented, transmitting informaiton about gender, age, attitude (friend or foe?), state of mind, and other aspects of the 'transmitter' of information. Human beings are also sensitive to detecting whether what is being said agrees with how it is being said, that is, to non-verbal information contained in the speech act, so. For example, we are able to detect 'intuitively' whether the speaker's intentions are honest, or whether he or she is lying.

Thus listening to a voice message can be done in two different ways: with a focus on linguistic information, leading to understanding it, and with a focus on speaker information, leading to recognize them and assess the way in which they act towards us.

Yesterday, Prof. Coleman used a quotation from Trubetzkoy to show that linguistics deals with whatever remains when all speaker-specific information is removed from the speech signal. Yet a converse of this claim is not necessarily true, as in speaker identification, we shall be looking for both those properties that are disregarded by a linguist, but we shall also deal with those aspects of phonetics and phonology that transmit speaker-related information.

A far-reaching goal of my work in recent months is to identify and model what cognitive processes underlie as simple an event as remembering one's voice and recognizing someone by voice.

MODES OF VOICE CHARACTERIZATION

Voice recognition of a known voice done by a human in every-day situations is not the only context in which this is done. It is only one extreme of a spectrum of implementation.

On a broader implementation plain, voice characterization requires the application of laboratory measurement with various degrees of involvement of automatic procedures (Machine-Aided Human Recogntion, Human-Aided Machine recognition). At the other extreme there are language-independent similarity judgements concerning e.g. two specimens of recorded voices done solely by the computer.

Let us now try and see what levels of linguistic analysis are involved in voice recognition by ear/brain.

PERSON RECOGNITION THROUGH THE AUDIO CHANNEL

Speech/voice characterization by ear/brain, performed simultaneously at numerous levels of perception

-**Phonetic**
-**Phonological**
-Phonostylistic
-Syntactic
-Stylistic
-Register-related
-Pragmatic
-Cognitive

OPERATIONAL UNDERSTANDING OF THE PHONETICS/PHONOLOGY DICHOTOMY

Phonetics/phonology: inseparable as all phonological phenomena are to be described by means of underlying phonetic mechanisms, but not vice versa. Phonology is the subject matter, while phonetics is the method of study.

Phonetics – studies **changes in the internal structure of sounds that do not affect their linguistic function** (phonemes/allophones) and remain below threshold values (eg. VOT not crossing the voiced-voiceless border). „Microphonetic". Such changes are difficult to perceive.

Phonological = **affecting articulation in a more conspicuous fashion.**

INTERACTION OF PHONETICS AND PHONOLOGY

Yet the fields of phonetics and phonology, albeit often regarded as inextricably interwoven, are a good example of disciplines based upon totally different methodological axioms, and can serve as a good example showing how employment of different principles can collaborate in reaching a reliable substantive conclusion.

Phonological analysis can be done by ear.
Phonetic analysis can be done in a laboratory,

CONTEXTS IN WHICH INDIVIDUAL FEATURES OF SPEECH ARE CRUCIAL:

- Pure research
- Engineering applications: devices whose action is triggered by oral commands:
  voice locks for VCE – voice controlled entry
  instantaneous identified command systems
- Forensic analysis: the need to identify a person in a recording. Speaker profiling (Kulsreshta, Singh and Sgharma 2012)

A typical scenario is that the police have an audio recording of an offender from a telephone intercept and another audio recording from an interview with a suspect.What the court wants to decide is whether the speaker on the two recordings is the same person or if the recordings come from two different people.The task of the forensic scientist is to analyze the acoustic properties of the voices on the recordings and on the basis of that analysis present a weight-of-evidence statement to help the court to make its decision. (Morrison: 2010)

VOCABULARY:
INDIVIDUAL SPEECH CHARACTERIZATION

KNOWN VOICES

Recognition – implies recognizing somebody we already know, typical situation in life. Ear/Brain

UNKNOWN VOICES

Comparison – S1 vs. S2, S1 vs. S3 … S1 vs. Sn – laboratory conditions

Identification - S1 vs. S2, S3, S4 … laboratory conditions,

Discrimination – Focus on decision: *This or not this speaker*. Ear/Brain, Laboratory conditions.

VOCABULARY

Voice Identification: phonetic parameters, language-independent

(Polish: identyfikacja głosu)

Speaker Identification: phonetic and phonological parameters; language-dependent

(Polish: Identyfikacja mówcy)

INDIVIDUAL SPEECH CHARACTERIZATION FEATURES - SUBJECT MATTER - METHOD

SUBJECT MATER	METHOD OF STUDY:	
	AURAL	LABORATORY
Non-linguistic individual spectral features	YES	YES
Phonation characteristics	YES/NO	YES
Durational characteristics	YES/NO	YES
Phonetic sub-allophonic qualitative variability	YES/NO	YES
Phonological features allophonic/phonemic variability	YES	YES

INDIVIDUAL SPEECH CHARACTERIZATION FEATURES OF A BROADER CONTEXT

BROADER CONTEXT	VALUES	
Emotional foundation	Emotionless	Strong emotional load
Phonostylistic variability	Normal pace	Fast/Slow pace
Accentual characterization	Marked	Unmarked
Health/anatomy related	Normal	Pathological
Stylistic/Register-related	Natural	Unnatural
Pragmatic	Matching the context	Not matching the context
Cognitive	Highly developed	Basic

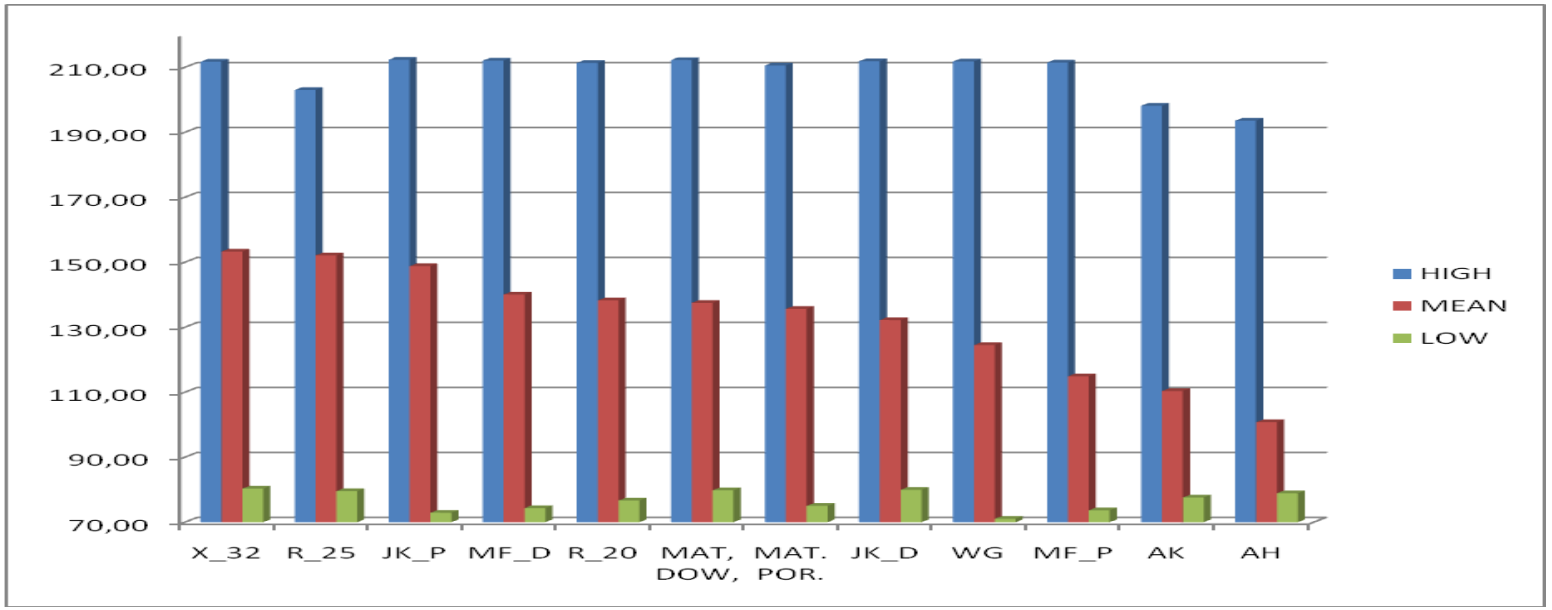
INDIVIDUAL SPEECH CHARACTERIZATION FEATURES OF A STILL BROADER CONTEXT

CONDITIONS	**VALUES**	
Distance/Direction	Stable	Unstable
Ambient noise	High SNR	Low SNR
Interaction	Clean	Overlaping interactions
Naturalness of speech	Normal	Deformed, impersonated
Recording device (hardware)	Good external microphone	Poor internal microphone
Recording device - software	Appropriate codecs	Ill-matched codecs
Transmission chanel	Broad range	Limited range

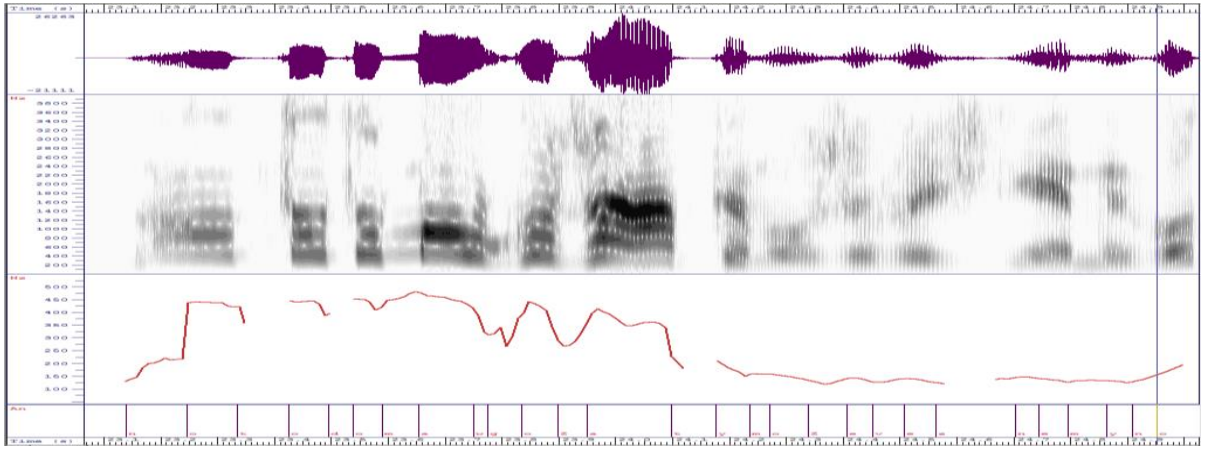
FACTORS TAKEN INTO ACCOUNT WHEN IDENTIFYING A PERSON vs. METHOD OF STUDY

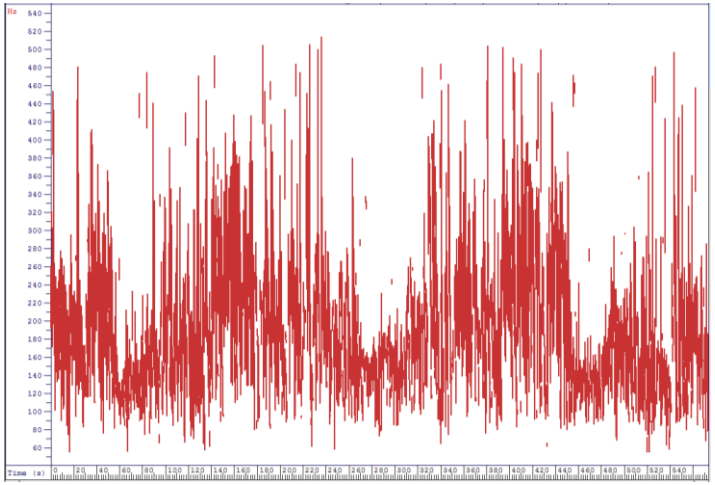
- **Aural / Phonetic information / Laboratory description**
  - **general impression of identity**
  - **voice characteristics** (pitch range, intensity, volume, rate)
  - **speech quality** (rhythm, fluency, pacing, phrasing and blending (Raje 2013)
  - **prosody** (duration, rhythm, fluency, pacing (Zeterholm 1997)
  - **intonation pattern**
  - **speech style**
  - **accent/dialect**
  - **emotional state**
  - **speech abnoramlities** (including speech pathology)
  - **speech naturalness** (natural, unnatural, distorted, impersonated)
  - **transmission chanel characteristics**

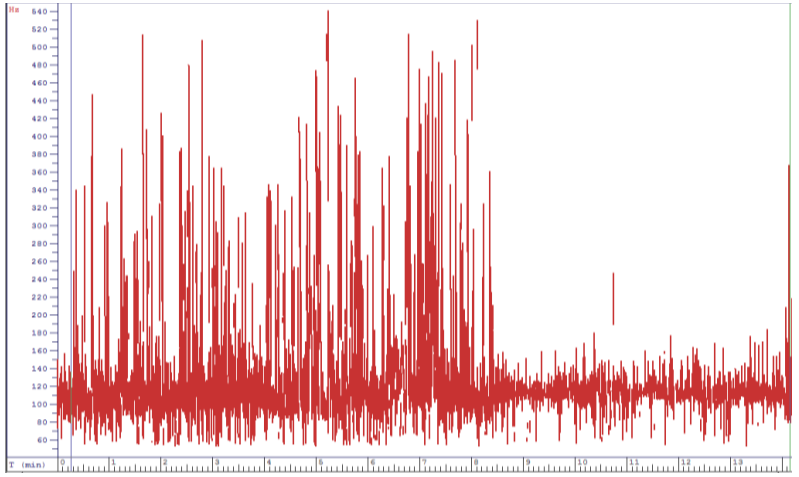
Example: Fo/Intonation. Database background



Fo/INTONATION – falsetto 545 Hz







FUNDAMENTAL FREQUENCY

A word of caution: results of measurements heavily depend on software settings!

Distorting emotions

Fragment of a recording

Ja go nie uderzył ani razu

EMOTIONS VS. PITCH

	EVIDENCE	PERSONAL SPECIMEN
Median pitch:	162.072	158.813
Mean pitch:	160.414	157.158
Standard deviation of period:	30.887	35.440
Minimum pitch:	73.513	65.790
Maximum pitch:	233.079	233.683

FACTORS TAKEN INTO ACCOUNT WHEN IDENTIFYING A PERSON vs. METHOD OF STUDY

- **Aural / Phonological information**

A: Z tego dżewa co tu może, wieź tam… wyżej na przykła.
B: Ja, ja.
A: Czyli na dziś…
B: No to zrób ta jag leci.
A: Ja?
B: Ja. Bardzo dobrze.
A. Kidy to mogę podjechać? Dożmi znać.
B. Dom ci znać. Bo teraz… deski to… u mnie jest ostateczno[ź], nie, corna.

A: Z tego dżewa co tu mo**ż**, wie**ź** tam… wy**ż**ej na przykła.
B: Ja, ja.
A: Czyli na dzi**ś**…
B: No to zrób ta jag leci.
A: Ja?
B: Ja. Bardzo dobrze.
A. Kiidy to mogę podjecha**ć**? Do**ż** mi zna**ć**.
B. Dom ci zna**ć**. Bo teraz… deski to… u mnie jest ostateczno[**ź**], nie, corna.

WORD-FINAL PRE-SONORANT VOICING: mosz -> mo**ż**, jak -> jag, dasz - do**ż**
FINAL CONSONANT DELETION: przykład -> przykła, tak -> ta,
VOWEL SHIFT: kiedy -> kiidy, dasz – do**ż**, masz – mo**ż**, dam -> dom, czarna -> corna
SPIRANT SHIFT: czarna -> corna
CLUSTER SIMPLIFICATION (+VOICING): drzewa -> dżewa
ostateczno**ś**ć -> ostateczno**ź**

DIALECT AND DIALECT SHIFT

(Sjöström et al. (2006). A Switch of Dialect as Disguise.

The attribute dialect is of high importance in the identification process. Listeners find it much more difficult to identify the target voice when a shift of dialect in the voice takes place. One possible reason for the results is that when making judgments about a person's identity, dialect as an attribute is strong and has a higher priority than other features.

A switch of dialect can easily fool listeners. This undermines earwitness identification of dialect and suggests that forensic practitioners who currently use dialect as a primary feature during analysis, need to reduce their reliance on this feature and be aware that they can easily be misled.

FACTORS TAKEN INTO ACCOUNT WHEN IDENTIFYING A PERSON vs. METHOD OF STUDY

Laboratory /phonetic information

Phonetic: measurements of selected sound parameters:

VOWELS: formant values, formant transitions (duration and direction), sound duration, onset/steady state/offset proportions, attack/sustain/decay dynamic characteristics, transitions in certain contexts, degree of nasalization before nasal consonants
CONSONANTS: VOT, VIC, COG, Spectral moments (Rodman et al. 2007)

Laboratory /phonological information

Phonological factors:

Phoneme/allophone switch, degree of palatalization, (e.g. palatal fronting), precision of articulation/degree of simplification (consonantal clusters),

VOWEL FORMANTS

Formant frequencies can be measured either statically, e.g. only at the centre of the sound, or dynamically, to investigate the dynamics of formant frequencies, which reflect the movement of a person's speech organs and are likely to reveal more fine-grained differences among speakers.

Formant frequency dynamics carry considerable speaker-specific information. By taking measurements along the formant contours, a significant improvement in speaker discrimination is achieved.

VOWEL FORMANT MEASUREMENTS – PROBLEM 1

Harrison 2006 studied a large body of data and showed that the formant measurements varied both within and between different software programs currently used in the field of forensic phonetics (3 programs – Praat, Multispeech & Wavesurfer) and between 3 analysis parameters (LPC order, analysis (frame/window) width, pre-emphasis).

Formant measurements are influenced by several factors including the method of analysis used. The overall degree of variation is different for each of the analysis parameters. The largest variation occurs when the LPC order is varied. It is difficult to make an overall judgement between the pre-emphasis and the frame width settings as to which produces the least variation.

The consequence of the study for forensic phoneticians is that they should be aware of the differences that altering analysis settings can have on formant measurements.

VOWEL FORMANT MEASUREMENTS – PROBLEM NO. 2A

Catherine Byrne and Paul Foulkes*
University of Birmingham Press 2004
Speech, Language and the Law 11(1) 2004
The 'Mobile Phone Effect' on Vowel Formants

Nolan et al. (2008). Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence
investigated the extent to which telephone transmission affects a listener's ability to distinguish among similar-sounding voices,

Both studies have shown that there exists a strong effect of mobile phone transmission on vowel formant frequencies.

VOWEL FORMANT MEASUREMENTS – PROBLEM NO. 2B

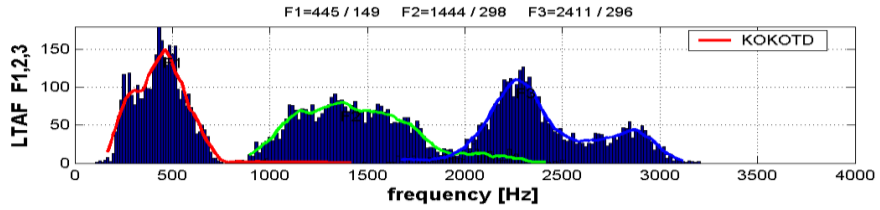
Guillemin and Wilson studied the impact of the GSM AMR Speech Codec on Formant Information

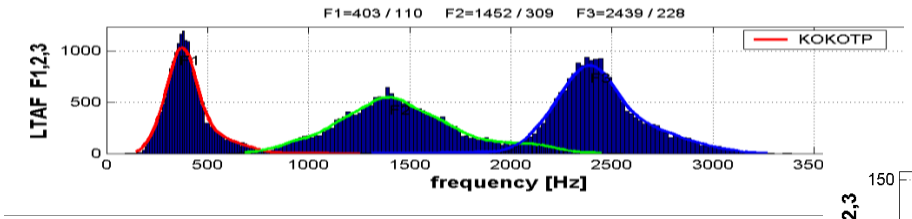
The Adaptive Multi-Rate (AMR) codec can be set at its various bit rates, and the settings exert an effect on acoustic parameters in the speech signal important for the task of forensic speaker identification (FSI).

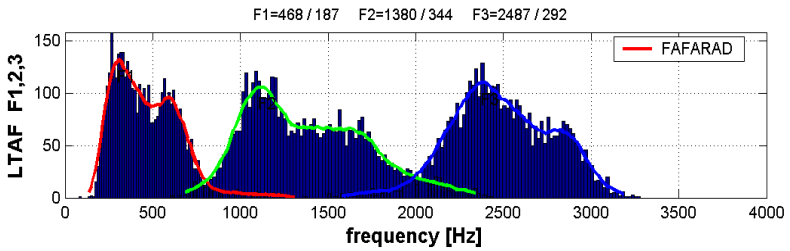
The acoustic parameters that are affected are the first three formant frequencies. It is shown that though the impact on these parameters as a function of bit rate can be quite significant, there is no consistent trend.

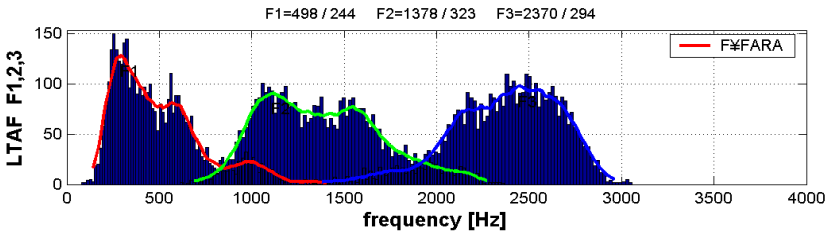
However, there are clear gender differences, likely caused by differences in pitch, with higher pitch female speech being affected significantly more by the codec than that of lower pitch male speech. In general formant frequencies are decreased by the codec, particularly in the case of high-frequency formants. These findings are significant to the FSI task and sound a distinct note of caution when analyzing speech that has been transmitted over the cell phone network utilizing this particular codec.

VOWEL FORMANT MEASUREMENTS – SIMILAR CHANNELS

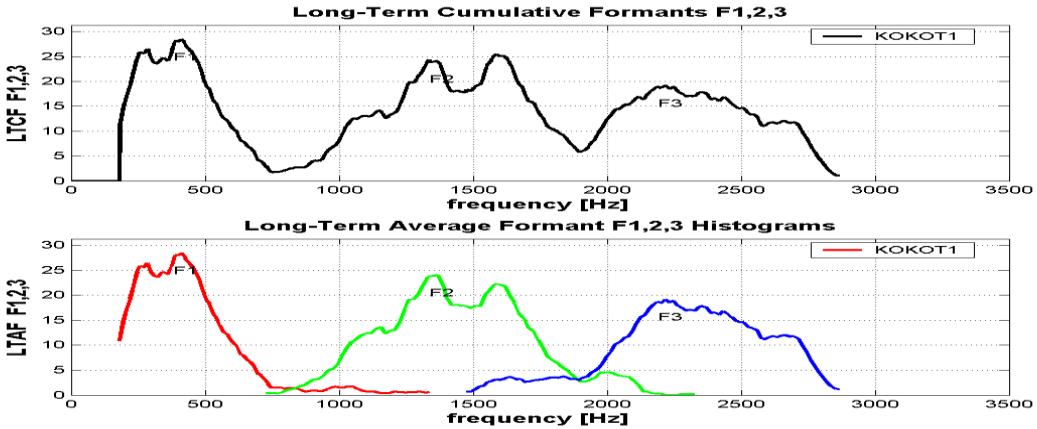


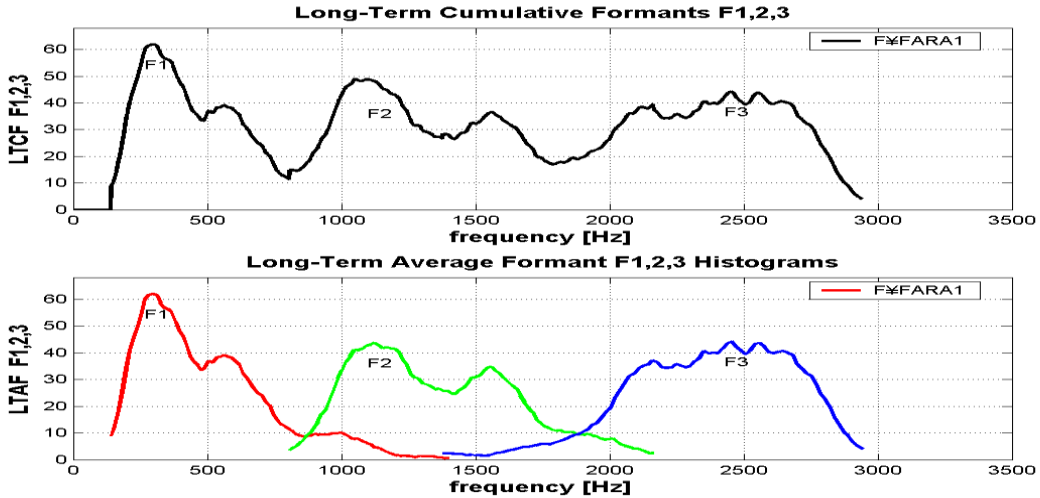




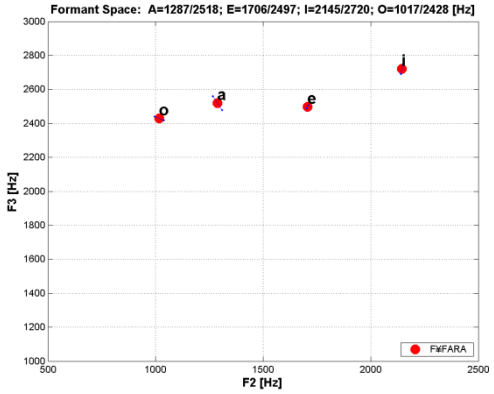


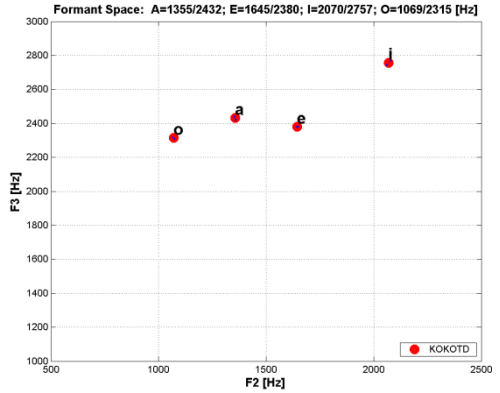
VOWEL FORMANT MEASUREMENTS – SIMILAR CHANNELS

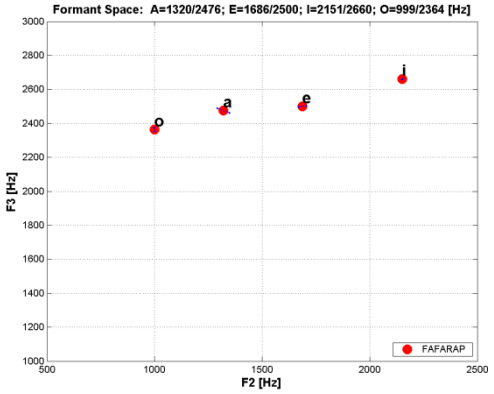


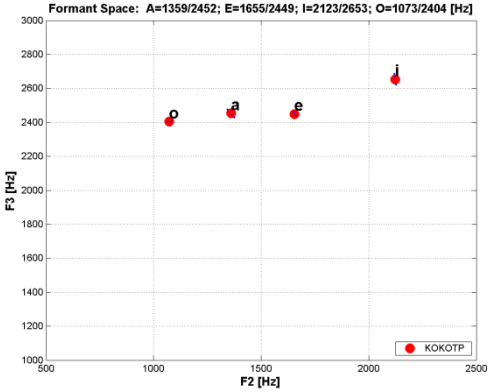


VOWEL PLOTS – SIMILAR CHANNELS









EMPIRICAL DATA

830 pairs of utterances by the same persons in different conditions.

FACTORS CRUCIAL FOR IDENTIFICATION (PERCEPTON OF SIMILARITY):
general impression of identity
pitch range
rate of speaking
intensity
fluency
emotional state
accent/dialect
speech abnormalities/pathologies
transmission channel characteristics
formant measurements
precision of articulation/degree of simplification

FACTORS THAT DO NOT EXERT EFFECT ON IDENTIFICATION
sound duration
degree of nasalization
vowel formant bends

PRELIMINARY CONCLUSIONS

In the context of FSI, the distinction between phonetic and phonological features is possible to define on the basis of the differences in operational methods of implementation.

A strong effect on unknown speaker characterization was exerted by non-phonetic non-phonological criterion, i.e., channel characteristics. For known voices, transmission chanel characteristics exerted a weak or no effect.

Fo characteristics can substantially aid FSI

Formant measurementsa should be implemented with caution

Phonological features (accent/dialec, cluster simplification, vowel shifts) are a good tool of characterizing a speaker.

The emotional state of the speaker can hinder the process of FSI

FUTURE QUESTIONS: SIGNIFICANCE OF FACTORS

What features carry strongest individual information?

What features override other features and distort perception of identity?

Is there a chierarchy of their strength?

What is the intra-speaker and inter-speaker variability?

STATE OF THE ART

In 2003, Jean-François Bonastre,[1,3] Frédéric Bimbot,[1,4] Louis-Jean Boë,[1,5] Joseph P. Campbell,[2,6*] Douglas A. Reynolds,[2,6*] Ivan Magrin-Chagnolleau[2,7]

Person Authentication by Voice: A Need for Caution

Our main conclusion is that, despite the existence of technological solutions to some constrained applications, at the present time, there is no scientific process that enables one to uniquely characterize a person's voice or to identify with absolute certainty an individual from his or her voice.

IMPORTANT AREAS NOT MENTIONED IN THE PRESENTATION

1. Automatic Speaker Identification Systems
2. Statistical background of the decision-making process
3. The use of databases in FSI

There is a need for a Unified Multi-Level System for Speaker Identification (forthcoming).

REFERENCES

Bonastre, J. F., Bimbot, F., Boë, L. J., Campbell, J. P., Reynolds, D. A., & Magrin-Chagnolleau, I. (2003). Person authentication by voice: A need for caution. In *Eurospeech 2003 - Interspeech 2003. Proceedings of the 8th European conference on speech communication and technology* . (pp. 33-6). Geneva, Switzerland. 1-4 September, 2003. Retrieved from http://www.afcp-parole.org/doc/AFCP_SpLC_HotTopicsEurospeech03_final.pdf

Byrne, C, and Folkes, P. (2004) University of Birmingham Press 350-1771*Speech, Language and the Law 11(1)*

Guillemin, B. J., C. I Watson (2011). Impact of the GSM AMR Speech Codec on Formant Information Important to Forensic Speaker Identification. Auckland: The University of Auckland

Harrison P. (2004) VARIABILITY OF FORMANT MEASUREMENTS
A thesis Submitted in partial fulfilment of the degree
of MA at the department of Language and Linguistic Science, University of York

Kulsreshta, S, and Sgharma A. (2012), In: Neustein and Hemant (eds.) Forensic Speakers Recognition. Law Enforcement and Counter-Terroriasma. Springer: Fort Lee.

Morrison, G. S. (2010) A Revolution in Forensic Voice Comparison. 2nd Pan-American/Iberian Meeting on Acoustics

Nolan, F., K. McDougall, K. and T. Hudson (2008) . Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence.
file:///E:/5.%20Literatura/LITERATURE/01.%20IDENT,%20VERIF,%20COMPAR,%20DISCR,%20RECOGN/0%200%200%20SELECTED/voice_similarity.html

Raje, N. (2013). **Forensic Speaker Identification & Speaker Profiling by Employing Forensic Phonetics, Aural-Acoustic Method & Identifying Abnormal & Pathological Speech.**
file:///E:/5.%20Literatura/LITERATURE/01.%20IDENT,%20VERIF,%20COMPAR,%20DISCR,%20RECOGN/0%200%200%20SELECTED/forensic-speaker-identification-speaker-profiling-by-employing-forensic-phonetics-auralacoustic-method-identifying-abnormal-pathological-speech.html

Rodman, D. McAllister,* D. Bitzer,* L. Cepeda* and P. Abbitt†*
Forensic speaker identification based on spectral moments
**Voice I/O Group: Multimedia Laboratory, Department of Computer Science*
North Carolina State University.

Sjöström, M. E.J. Eriksson, E. Zetterholm, and K.P. H. Sullivan. A Switch of Dialect as Disguise. *Working Papers 52 (2006), 113–116.* Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics

TRAWIŃSKA, A i M. KAJSTURA. (2004). **THE INBUILT RECORDER OF MOBILE PHONES . POSSIBILITIES OF FORENSIC SPEAKER IDENTIFICATION** *Problems of Forensic Sciences, vol. LVII, 2004, 51.80*

Zetterholm, E. **The role of prosody in voice imitation,** Lund University, Dept. of Linguistics 269 *Working Papers 46 (1997), 269–287*

THANK YOU!