Co-funded by the Internal Security Fund of the European Union

Best Practice Manual

for the Methodology of Forensic Speaker Comparison

ENFSI-FSA-BPM-003

Version 01 – December 2022

ENFSI's position on Best Practice Manuals

ENFSI wishes to promote the improvement of mutual trust by encouraging forensic harmonisation through the development and use of Best Practice Manuals. Furthermore, ENFSI encourages sharing Best Practice Manuals with the whole Forensic Science Community which also includes non ENFSI Members.

Visit www.enfsi.eu/documents/bylaws for more information. It includes the ENFSI policy document Policy on Creation of Best Practice Manuals within ENFSI (code: QCC-BPM-001).

European Union's Internal Security Fund – Police
This Best Practice Manual for the Methodology of Forensic Speaker Comparison was funded by the European Union's Internal Security Fund – Police.

Acknowledgements
The realisation of this Best Practice Manual was mainly driven by forensic experts of the ENFSI Forensic Speech and Audio Analysis Working Group (FSAAWG):

Isolde Wagner (Federal Criminal Police Office of Germany, Forensic Science Institute, Germany),
Dr. Dagmar Boss (Bavarian State Criminal Police Office, Forensic Science Institute, Germany),
Dr. Vincent Hughes (Department of Language and Linguistic Science, University of York, UK),
Timur Svirava,
Ivan Siparov,
Dr. Miriam Rolfes (Berlin State Criminal Police Office, Forensic Science Institute, Germany).

Many thanks to all colleagues and laboratories who gave their time to provide comments as part of the public review.

Special thanks to Dr. Michael Jessen (Federal Criminal Police Office of Germany, Forensic Science Institute, Germany) for collecting a huge amount of valuable literature on Forensic Speaker Comparison. Many entries of this Bibliography were adopted from his book (2012).

Official language
The text may be translated into other languages as required. The English language version remains the definitive version.

Further information
For further information about this publication, contact the ENFSI Secretariat. Please check the ENFSI website (www.enfsi.eu) for updated information.

ISBN Identifier
978-3-949949-05-0

BEST PRACTICE MANUAL FOR THE METHODOLOGY OF FORENSIC SPEAKER COMPARISON

DOCUMENT TYPE:	REF. CODE:	ISSUE NO:	ISSUE DATE:
BPM	**FSA-BPM-003**	**001**	**12.12.2022**

TABLE OF CONTENTS

1. AIMS

This Best Practice Manual (BPM) aims to provide a framework for procedures, quality principles, training processes and approaches to forensic examination. This BPM is aimed at experts in the field of Forensic Speaker Comparison (FSC) and assumes prior knowledge and qualifications in the discipline. The BPM can be used by member laboratories of ENFSI and other forensic science laboratories to establish and maintain working practices in this field, to deliver reliable results, maximise the quality of the information obtained and produce robust results. Literature with more detailed information on different aspects of the methodology of FSC is provided in Appendix 2 ('Bibliography'). The use of a consistent methodology and the production of comparable results will facilitate interchange of data between laboratories.

The term BPM is used to reflect the scientifically accepted practices at the time of writing. The term BPM does not imply that the practices laid out in this manual are the only good practices used in the field.

2. SCOPE

FSC involves the analysis of audio recordings containing the voices of unknown and known speakers in order to help answer the question of whether these voices belong to the same speaker or different speakers. The methodology described in this BPM is based on a combined procedure of phonetic-linguistic auditory and acoustic analyses of different speech features.

This BPM does not address automatic and semiautomatic speaker recognition. For more information see the ENFSI document 'Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition' [01].

It is important to acknowledge that different methodologies have advantages and disadvantages. It is possible to combine automatic, semi-automatic and linguistic-phonetic-acoustic methodologies. The use of different methodologies can provide complementary and thus more information. However, the specific choice of methodologies is dependent on the suitability of the material (in terms of quality, quantity etc.) as well as on the availability of methodological techniques within a given laboratory and the rules of the jurisdiction in which the expert works.

3. TERMS AND DEFINITIONS

Explanations are included in the text. More details can be found in the literature provided in Appendix 2 ('Bibliography').

4. RESOURCES

4.1 Personnel

The methodology of FSC requires experts in the field of forensic phonetics and linguistics or related sciences in speech and audio analysis. The expert is responsible for the application of this methodology to a given case and has the key role of directing and performing the examination. The examination usually includes auditory and acoustic analyses of the speech material, interpreting findings, providing the results, writing the expert report and presentation at court.

4.1.1 Expert Qualification

A master's degree (or equivalent) in phonetics and linguistics or related sciences in speech and audio analysis is strongly recommended as a minimum requirement.

4.1.2 Training and Assessment

Laboratories should have written requirements of competence for their experts. There should be a documented training program and processes for assessing whether trainees have achieved the required level.

There should be initial training before experts are authorised to undertake casework. The expert should also be subjected to regular assessment to ensure that competence is maintained and developed. This can be accomplished through a number of different mechanisms including formal tests, undeclared trials, and peer review of case analysis and reports.

In addition to any formal assessment by their organisations, experts should read professional and academic literature containing pertinent information relating to forensic phonetics and FSC specifically, take part in appropriate events, e.g. workshops, seminars, training courses, etc., and whenever possible, actively participate in research and development projects.

4.2 Equipment

The equipment regularly used in FSC consists of computer devices with specific software for acoustic play-back and signal processing as well as various acoustic devices, such as different kinds of loudspeakers, headphones, microphones, sound cards, amplifiers, etc.

In cases involving analogue recordings, appropriate analog-to-digital conversion (ADC) equipment must be used for digitisation.

The equipment inventory shall be documented and shall be shown to operate properly before being used in casework.

4.3 Reference Materials

Not applicable.

4.4 Facilities & Environmental Conditions

Forensic laboratories performing examinations in FSC shall be designed for efficient and effective operation.

As auditory and acoustic analyses are the key elements of the methodology, soundproofing is required. Environmental noises should not interfere with auditory and acoustic examinations within the laboratory. Preferably, there should be only one expert per room when conducting auditory analysis.

4.5 Risk-Based Thinking

FSC is a methodology which involves considerable subjectivity on behalf of the expert. Specific risks and opportunities shall be continuously identified, evaluated and ranked. The whole working process shall be regularly evaluated in terms of undesired implications and potential failures. Laboratories shall have determinations of actions to be taken and implementations of monitoring and follow ups of risks and opportunities.

4.6 Materials and Reagents

Not applicable.

5. METHODS

5.1 Principles

The methodology of FSC was developed for the purpose of comparing recordings of unknown against known speakers or different unknown speakers against each other.

If there is more than one sample of an unknown speaker within one case scenario to be analysed, the samples shall not be automatically pooled, i.e. assumed to originate from the same individual. They shall be analysed separately or at least checked in detail before being pooled.

A fundamental principle of FSC is that a range of speech features is analysed to capture the many dimensions on which speakers can be distinguished. Speech features should be as independent of each other as possible to reach a high degree of speaker-discriminatory power. The relationship between intra- and inter-individual speaker variation determines the relevant discriminatory information. All findings shall be compared and evaluated on the basis of the (dis-)similarity and the typicality of speaker-specific characteristics. In this process, the knowledge and competence of the expert plays a substantial role. The expert shall make decisions at every step of the FSC analysis using available scientific background information, as well as his or her experience. After the comparison and evaluation process a conclusion statement is given. The results of the examination are documented in an expert report.

The analysis process of this methodology combines auditory phonetic-linguistic perception and descriptions of speech features on the one hand, and acoustic measurements of the speech signal on the other hand.

5.2 Speech Feature Analysis

A wide range of discriminatory speech features could in principle be analysed within FSC. They can, for example, be categorised into segmental and supra-segmental features or can be put in parent categories like ‘language’, ‘voice’ or ‘manner of speaking’. Some can be analysed with both auditory and acoustic methods. The features chosen for analysis may differ from case to case, depending on what is available in the recordings and what is considered important by the expert. Speech features also differ in within-speaker variability (due to e.g. emotion, illness, drugs, alcohol). Some of the most commonly analysed features are described below.

5.2.1 Language, Dialect, and Foreign Accent

Features like language, dialect, and foreign accent in speech provide important information about a speaker’s origin, education, and further background. Analyses of those features help answer the question as to whether the language used was a native or a second language. Furthermore, such analyses can provide information about the type and degree of regional influence, and about the possible presence of a certain type of slang. Wording and pronunciation can provide specific social and/or individual feature characteristics.

The analyses cover both segmental and supra-segmental features. There is investigation into the articulation of vowels and consonants on the one hand, and prosodic patterns like intonation on the other hand. Additionally, there are linguistic analyses of lexis and grammar use in speech. The whole analysis process is principally auditory phonetic-linguistic oriented with acoustic analyses conducted where necessary and possible. This results in a detailed

description of speaker characteristics. Speaker characteristics may be described in relation to common use or standard speech, if such norms exist for the respective language.

If the expert is a non-native speaker of the language under analysis or does not have thorough knowledge of the language, he or she shall co-operate with a language consultant (e.g. linguist or translator) with appropriate qualifications.

5.2.2 Fundamental Frequency and Variation

Fundamental frequency (F0) is the frequency of the (quasi-)periodic structure of voiced speech signals. There are a number of F0 detection techniques, e.g. based on the cepstrum of the speech signal. F0 is usually measured in Hz and reported as a mean value. In addition to the mean, other F0 parameters can be reported, including median, max and min values, standard deviation and so on. F0 is the physical correlate of the perceived pitch.

Typical ranges of interspeaker variation for F0 mean values are 80-200 Hz for men and 150-400 Hz for women who normally have higher F0 due to shorter vocal fold length.

F0 can be influenced by the conditions under which the speech is being produced (noise, channel, emotions etc.) and thus cause intraspeaker variability. Factors like tension of the vocal cords (vocal folds), force of glottal closure, and expiratory air pressure lead to changes in F0. The mean values of F0 are reported to change slightly with age.

5.2.3 Voice Quality

The term voice quality refers to quasi-permanent characteristics that are specific for the overall sound of a speaker's voice. These characteristics are divided into two levels: phonation and resonance of the vocal tract.

Phonation refers to the glottal tone caused by the oscillation of the vocal folds in the larynx. This oscillation can be specific for a speaker's voice in terms of its short-term frequency and amplitude. Categories for the description of phonation are, for example, harsh, tense, breathy or creaky voice.

The source signal produced by phonation is then filtered by the vocal tract (consisting of the pharyngeal, oral and nasal cavities). The anatomic-morphologic conditions of the vocal tract and the individual neuro-muscular habitual patterns determine the resonance characteristics that are typical for each individual. Categories for the description of vocal tract influence are, e.g. nasality, lip rounding or spreading, and close or open jaw.

The characteristics of voice quality can be analysed both auditorily and where necessary and feasible by the evaluation of spectrograms or acoustic measurements (e.g. jitter, shimmer, etc.). The individual characteristics can be described in terms of frequency of appearance, degree, and representativeness. The analysis is conducted in terms of deviations from a hypothetical neutral voice quality, i.e. a voice that is unmarked in terms of the anatomy and physiology of voice production.

5.2.4 Formant Frequencies

Formants are the maxima of the vocal tract frequency response. Their position within the speech signal spectrum is determined both by the phonetics of the language or variety on the one hand and the individual shape of the speaker's vocal tract on the other. In most cases vocalic sounds are used for formant analysis.

Two commonly used approaches to formant analysis are:

- Single vowel formant analysis. It involves the collection of formant measurements from a specific vowel and its comparison across the speech samples under investigation.

- Analysis of long-term formant distribution (LTFD). It captures the distribution of values for each formant of a speaker over a given speech recording. LTFDs are frame-by-frame measurements – therefore, long vowels carry more weight than short vowels in that they yield a greater number of measurements per vowel. A positive attribute of LTFDs is that they do not require the categorisation of individual vowels into phoneme classes, as all vowels are considered in an analysis.

5.2.4.1 Channel Impact

As the speech signal is transmitted through different channels and digitally processed during the recording, the impact of these transformations should be considered during formant analyses. Microphone frequency response, telephone line and radio channel are examples of additional processes that can affect formant values. GSM (Global System for mobile communications) codecs are an example of digital processing that also affects the formants' positions as they store the signal's Linear Prediction Coefficients instead of raw audio data. Low bitrate MP3 compression stores spectral components according to the principles of the psychoacoustic model of hearing and deletes inaudible components. These and any other issues of this type must be considered during formant analyses.

5.2.5 Speech Tempo

Speech tempo is the speed at which an individual produces speech. It is generally analysed as either speaking rate or articulation rate. Speaking rate captures overall tempo across entire turns, which includes pauses, disfluencies, and non-linguistic articulations. Articulation rate is more narrowly defined as the rate of speech production within utterances, and therefore excludes pauses, disfluencies, and non-linguistic information.

While it is possible to judge qualitatively (using subjective labels like 'fast' or 'slow'), speech tempo is commonly analysed quantitively, measured as the number of syllables per second. When measuring tempo, the expert shall make a decision about whether to count the number of phonetic syllables (i.e. the number of syllables actually produced phonetically by the speaker) or the number of phonological syllables (i.e. the number of syllables in the underlying phonological representation).

Tempo varies across languages (as well as possibly dialect and other social factors, e.g. age). Speech tempo can also be affected by speaking style and situation and so displays considerable within-speaker variability.

5.2.6 Hesitation Phenomena and other Non-Pathological Speech Disfluencies

Hesitation phenomena are those features which either occur unintentionally or are used deliberately when a speaker hesitates or pauses before or within an utterance. There can be speaker discriminatory information in the duration, the frequency of occurrence, the voice quality, and other features such as an adjacent glottal stop.

Hesitation phenomena can be analysed in terms of the:

- Use of hesitation particles, such as 'uh/uhm' in English or 'äh/ähm' in German.
- Lengthening of vowels and consonants. Here it should be stated which vowels and/or consonants are affected most often.
- Occurrence of silent pauses. It should also be checked whether their use is related to the content of the utterance (e.g. emphasising the following part of the utterance) or if it is erratic.

Speech fluency can also be affected by disruptions of utterances. Some people tend to, e.g., break off utterances and restart them with a different wording, produce a completely new

utterance, or repeat parts of an utterance (sentences, single words, syllables and/or sounds – the last two sometimes are perceived as a kind of pathological stuttering).

Individual speakers tend to use a certain inventory of hesitation phenomena. The frequency of occurrence and combination of these phenomena can carry speaker-specific information but can also be affected by the speaking situation.

5.2.7 Speech Pathologies

The term speech pathology covers a wide range of communication and cognitive disorders. The underlying causes of speech pathologies are numerous and so a detailed discussion is beyond the scope of this document. However, such disorders can, in principle, affect speech production at any level of linguistic analysis. The most widely recognised of these is stuttering, an articulation disorder affecting the planning and implementation of speech production, which manifests as silences and/or repetition and prolongation of speech sounds. Stuttering may broadly be classified as a disfluency of which there are many kinds. However, pathology can also affect the phonetics of speech production, especially where an individual has some physical impairment (such as a cleft palate) or speech impediment (such as a lisp). In such cases, long-term features, either at the laryngeal (such as fundamental frequency or voice quality) or supra-laryngeal (such as habitual tongue body fronting) levels, are most likely to be affected.

Where they do present themselves, pathologies can be extremely useful markers of individual identity. In such cases, it may be helpful to engage with a speech and language therapist, who has greater experience and more detailed knowledge of how to analyse and categorise pathologies.

6. VALIDATION AND ESTIMATION OF UNCERTAINTY OF MEASUREMENT

6.1 Validation

The methodology of FSC has been developed within forensic laboratories for decades. The development has been conducted on the basis of scientific research in the field of forensic speech science and linguistics, through published peer-reviewed literature and empirical testing (under casework conditions), and through knowledge exchange within the community of researchers and experts (e.g. during conferences). The methodology consists of analysing individual speech features on different dimensions with appropriate methods, each of them having been developed and tested for its speaker discriminatory power in the aforementioned way (see the literature in Appendix 2 ‘Bibliography’). Individual analyses are then combined to arrive at an overall conclusion. New methods of feature analyses shall also be validated as they become available.

To validate the methodology of FSC as a whole, it shall be demonstrated that the methodology is able to separate same-speaker and different-speaker pairs with a given level of accuracy. This shall be demonstrated through suitable quality controls, like proficiency tests (PTs), collaborative exercises (CEs), and/or blind testing.

Additional guidance of validating laboratory procedures can be found in the ENFSI ’Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science’ [03].

6.2 Estimation of Uncertainty of Measurement

Not applicable.

7. QUALITY ASSURANCE

7.1 Proficiency Testing / Collaborative Exercises

Proficiency tests (PTs) and collaborative exercises (CEs) are necessary to test and to assure the quality of the methodology of FSC. The 'Guidance on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI' [02] provides general information for the ENFSI Expert Working Groups (EWGs) on how to organise effective PTs and CEs for their members.

Besides the general ENFSI guidance, there are some specific challenges in FSC. The test material is speech, which is language dependent, and thus varies through different countries and regions. Full feature analyses of the speech material (especially in respect of e.g. language and dialect) can only be made if the expert is a native speaker or has a thorough knowledge of the language involved. As test material must be consistent within a trial, PTs and CEs for the methodology of FSC are not organised on an international basis within ENFSI. They should be organised at a national or regional level. The organisation of such PTs and CEs and the frequency of participation should follow the rules of different ENFSI member laboratories and their accreditation bodies.

7.2 Quality Controls

Within the methodology of FSC, the procedure of feature analyses, evaluation of results, and reporting shall be documented with regard to the policy of local practices and whenever possible in Standard Operation Procedures (SOPs). The SOPs should be included in a training program, regularly reviewed, and updated if necessary.

When using the methodology of FSC for casework there should be sufficient documentation and/or clear analysis protocols for all feature analyses (including e.g. the selection of samples, details of duration and acoustic quality or measurement values, etc.), such that another expert would be able to understand and replicate the results.

Since the methodology of FSC is human-based to a high degree and dependent on the competence and knowledge of the expert, steps shall be taken to mitigate the effects of cognitive bias and to avoid nonconforming work. Both could lead to false results and misinterpretation.

7.3 Data Collection for Control, Monitoring and Trend Analysis

Not applicable.

7.4 Verification / Peer Review

The findings of an FSC analysis and the final report should be carefully checked by a second expert as part of peer review (including blind peer review) if it is possible and admitted by local legislation. All peer reviews shall be documented in the case file.

8. HANDLING ITEMS

8.1 At the Scene

Not applicable.

8.2 In the Laboratory

The material to investigate is an acoustic signal which can reach the laboratory in different ways and formats.

The submitted material shall be registered. It shall be checked to assess whether material matches the description in the request and if it is complete. If needed, prioritisation according to the laboratory's regulations and the customer's request shall be considered. Confidentiality of the case material shall also be considered. The customer shall be consulted to clarify possible questions about the material. Each step of handling items and the working procedure shall be documented.

8.2.1 Digital Data

In most cases the material is delivered in a digital format either via data transfer or on a digital carrier. The original material should be stored as a 1-to-1 copy with write protection before any processing of the data (e.g. channel separation, conversion, editing etc.).

8.2.2 Analogue Data

In rare cases, the material can be submitted on an analogue medium, like a wheel tape or an audio cassette. In such cases, the audio signal shall be digitised in an appropriate way. A suitable device shall be chosen and checked with a test tape to assess whether it functions correctly. Information loss and/or corruption by low sampling rate, low-bit quantisation and lossy file formats shall be avoided. A minimum sampling rate of 44.1 kHz, 16-bit quantisation and PCM (Pulse-Code-Modulation) Wave target format are appropriate for most digitisations. For the speech feature analyses themselves, only the digital copy shall be used.

9. INITIAL ASSESSMENT

In the initial assessment all relevant characteristics of the audio recordings of a given case shall be examined, including duration and various technical and contextual aspects of speech. Since specific characteristics of the material may have an impact on the analyses and their outcome within FSC, the knowledge and experience of the expert is necessary to check whether the material can be analysed and, if so, what kind of feature analyses may be applicable. All determining factors shall be checked and evaluated.

9.1 Quantity of the Forensic Audio Material

One important factor is the quantity of the forensic audio material. This can be examined by both measuring the net duration of the speech samples and analysing their representativeness in terms of the amount of individual verbal behavior included in the samples. In principle, recordings of less than about 10 seconds are not expected to contain much speaker-specific information. But even longer samples can sometimes provide only little speaker-specific information.

9.2 Acoustic Quality of the Forensic Audio Material

Another important factor that shall be checked is the acoustic quality. Forensic audio material often comes from transferred (telephone) recordings with reduced technical properties, including filter effects, reduced signal to noise ratio, environmental noise, compression and lossy formats, delay effects or other distortions and artefacts. Greatly reduced acoustic quality can make feature analyses difficult or even not applicable.

9.3 Contextual Aspects

In addition to the quantity and quality aspects of the audio material there are a variety of contextual aspects that can have further limitations to features analyses within FSC. These include increased vocal loudness, stress and emotion, speaking with reduced intelligibility, speaking while fatigued, speaking under the influence of intoxication, etc.

Furthermore, and specific to the forensic context, is the limitation through voice disguise. Voice disguise can occur through behavioural modifications (e.g. speaking in falsetto voice, using a mouth cover, imitating a foreign accent) or through the use of technical equipment (e.g. voice changer, deep fakes).

9.4 Mismatched Conditions between Questioned and Reference Recordings

A factor that commonly occurs in FSC is the situation of mismatched conditions between questioned and reference samples. Mismatch is defined here as any kind of different characteristics between recordings, e.g. telephone vs. microphone recordings, recordings with background noise vs. in a quiet room, read vs. spontaneous speech, loud vs. neutral voice, or non-contemporaneous speech. There are a variety of possibilities.

The influence of any mismatched condition on feature analyses shall be checked carefully and taken into account in the whole investigation of FSC. Higher caution shall be exercised in analysing cases involving language mismatch where phonetic-linguistic feature characteristics of speech can differ completely between different languages.

If the reference material is not provided by the customer recordings can be made by the expert. In this case the recommendations in Appendix 1 ‘General Remarks about Reference Recordings for FSC’ should be followed.

10. PRIORITISATION AND SEQUENCE OF EXAMINATIONS

The expert is responsible for prioritisation and the sequence of examinations.

11. RECONSTRUCTION

Not applicable.

12. ASSESSMENT OF RESULTS AND INTERPRETATION

After detailed analyses of various speech features (the most common of which are described in section 5), all similarities and differences observed between the compared speakers shall be evaluated by the expert. The evaluation process shall include considerations about the degree of suitability and comparability of the material as well as considerations about the characteristics of features and their frequency of occurrence in relation to inter- and intra-individual variation. The interpretation shall be carried out both by evaluating the results of single feature analyses at every step and in combination with each other.

Conclusions in FSC cases are in many respects subjective and based to a large extent upon the competence (education, training and experience) of the expert. Empirical and statistical analyses remain challenging for many speech features due to the lack of background data (e.g. in the context of language, dialect etc.).

At present, there is no universally used scaling for reporting conclusions in FSC cases. The type and range of scales differ widely between different laboratories and jurisdictions. Statements shall be considered and expressed with thoroughness and care.

A framework for evaluation and interpretation is described in the ‘ENFSI Guideline for Evaluative Reporting in Forensic Science’ [04].

13. PRESENTATION OF RESULTS

Depending on legal regulations, results can be presented to the court orally and/or in written form. The presentation shall clearly state the results of any evaluation and interpretation of the examination.

Written reports shall include all the relevant information in a clear, concise, structured and unambiguous manner as required by the jurisdiction. Reporting shall also be in accordance with the existing laboratory regulations.

14. HEALTH AND SAFETY

During the process of auditory analysis, it must be taken into account that both increased long-term exposure to audio signals with high sound pressure levels and especially single loud events like pulses can cause psychological stress, tinnitus, and hearing impairment up to hearing loss.

Listening over an extended time span can face similar health risks. For this reason, pausing is required depending on the equipment that is used. Experience has shown that, when listening to loudspeakers, there should be a minimum of 30 minutes break after at most two hours of intense auditory analysis. When using headphones, intervals of uninterrupted listening should be reduced to one hour.

Furthermore, work in the forensic field, especially when analysing audio or video recordings with distressing content such as murder or child abuse can cause psychological problems like persistent stress, anxiety, and trauma. In such situations, psychological support should be sought.

15. REFERENCES

[1] Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J. & Niemi, T. (2015) *Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition*. Frankfurt: Verlag für Polizeiwissenschaft. http://enfsi.eu/wp-content/uploads/2016/09/guidelines_fasr_and_fsasr_0.pdf

[2] QCC-PT-001, Guidance on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI, version 001, 27/06/2014.

[3] QCC-VAL-002, Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science, version 001, 10/11/2014.

[4] ENFSI guideline for evaluative reporting in forensic science: Strengthening the Evaluation of Forensic Results across Europe, version 3.0, 08/03/2015. https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

16. AMENDMENTS TO PREVIOUS VERSION

Not applicable (first version).

APPENDIX 1

General Remarks about Reference Recordings for FSC

For the recording of reference speech samples, it is preferable to use a digital recording device. The recording should be made in an uncompressed format.

The reference recording should allow for auditory, linguistic, and instrumental analyses. Therefore, appropriate acoustic quality is required. The recording should be made in a room with a low level of reverberation. All extraneous noises and interference (e.g. knocking, slamming the door, whispering, other voices, typing, phone calls, etc.) shall be minimised.

Since different situational and technical conditions of recordings can influence the comparability of speech samples, it is desirable to achieve the greatest obtainable match between the reference recording and the existing questioned audio. Technical characteristics of the reference recording should at least be equivalent if not better than those of the questioned audio. Steps shall be taken to avoid possible voice disguise of the recorded person.

To assure sufficient representativeness of speaker-individual behaviour within the reference material net speech duration of the recording should exceed 5 minutes in total.

The reference recording should be saved on a data storage device with write protection and possibly encrypted. If not transmitted via data transfer the material should be properly packaged before submitting to the forensic speech and audio laboratory.

APPENDIX 2

Bibliography

The following list contains a collection of publications that relate to the traditional methodology of Forensic Speaker Comparison (forensic phonetic-linguistic auditory and acoustic approach).

Allen, J.S., Miller, J.L. & DeSteno D. (2003) Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113: 544-552.

Huntley Bahr, R. & Frisch, S.A. (2002) The problem of codeswitching in voice identification. In Braun & Masthoff (eds.) *Phonetics and its applications. Festschrift for Jens Peter Köster on the Occasion of his 60th Birthday.* Stuttgart: Steiner. 86-96.

Baldwin, J. & French, P. (1990) *Forensic Phonetics*. London: Pinter.

Batliner, A., Kießling, A., Burger, S. & Nöth, E. (1995) Filled pauses in spontaneous speech. *Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm).* 472-475.

Beckman, M.E. & Venditti, J.J. (2010) Tone and intonation. In: Hardcastle, Laver & Gibbon (eds.) *The Handbook of Phonetic Sciences* [2nd ed]. Chichester: Wiley-Blackwell. 603-652.

Böhm, T., Uiváry, I. & Németh, G. (2009) The occurrence of irregular phonation as a speaker characteristic in Hungarian. *The Phonetician* 99/100: 45-58.

Bóna, J. (2014) Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America* 136(2): 116-121. http://dx.doi.org/10.1121/1.4885482

Boss, D. (1996) The problem of F0 and real-life speaker identification: a case study. *Forensic Linguistics* 3: 155-159.

Bradlow, A.R., Ackermann, L., Burchfield, L.A., Hesterberg, L., Lique, J. & Mok, K. (2011) Language- and talker-dependent variation in global features of native and non-native speech. *Proceedings of the International Congress of Phonetic Sciences 17 (Hong Kong)*. 356-359.

Braun, A. (1995) Fundamental frequency – How speaker specific is it. In: Braun & Köster (eds.) *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag Trier. 9-23.

Braun, A. (1996) Age estimation by different listener groups. *Forensic Linguistics* 3: 65-73.

Braun, A. & Rosin, A. (2015) On the speaker-specificity of hesitation markers. *Proceedings of the International Congress of Phonetic Sciences (Glasgow)*. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0731.pdf

Braun, A. & Wagner, A. (2002) Is voice quality language dependent? In: Braun & Masthoff (eds.) *Phonetics and its applications. Festschrift for Jens Peter Köster on the Occasion of his 60th Birthday.* Stuttgart: Steiner. 298-312.

Braun, A. & Friebis, S. (2009) Phonetic cues to speaker age: A longitudinal study. In Grewendorf & Rathert (eds.) *Formal Linguistics and Law*. Berlin: Mouton de Gruyter. 141-162.

Byrne, C. & Foulkes, P. (2004) The 'mobile phone effect' on vowel formants. *The International Journal of Speech, Language and the Law* 11: 83-102.

Cambier-Langeveld, T. (2007) Current methods in forensic speaker identification: Results of a collaborative exercise. *The International Journal of Speech, Language and the Law* 14: 223-243.

Cambier-Langeveld, T., van Rossum, M. & Vermeulen, J. (2014) Whose voice is that? Challenges in forensic phonetics. In: Caspers, Heeren, Pacilly, Schiller & van Zanten (eds.)

Above and Beyond the Segments. Experimental linguistics and phonetics. Amsterdam: John Benjamins Publishing Company. 14-27.

Cao, H. & Wang, Y. (2001) A forensic aspect of articulation rate variation in Chinese. *Proceedings of the International Congress of Phonetic Sciences 17 (Hong Kong)*. 396-399.

Castellana, A., Carullo, A., Astolfi, A., Puglisi, G.E. & Fugiglando, U. (2017) Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. Journal of the Acoustical Society of America 141(4): 2353-2363. http://dx.doi.org/10.1121/1.4979115

Chen, N., Shen, W., Campbell, J. & Schwartz, R. (2009) Large-scale analysis of formant frequency estimation variability in conversational telephone speech. *Proceedings of INTERSPEECH (Brighton)*. 2203-2206.

Chin, S. & Pisoni, D. (1997) *Alcohol and Speech*. San Diego: Academic Press.

Clark, H. & Tree, J. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition* 84: 73-111.

Clark, J. & Foulkes, P. (2007) Identification of voices in electronically disguised speech. *The International Journal of Speech, Language and the Law* 14: 195-221.

Corley, M. & Stewart, O. (2008) Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass* 2: 589-602.

Dankovičová, J. (1997) The domain of articulation rate in Czech. *Journal of Phonetics* 25: 287-312.

De Boer, M. & Heeren, W. (2019) The speaker-specificity of filled pauses: A cross-linguistic study. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 607-611.

De Castro, A., Ramos, D. & Gonzalez-Rodriguez, J. (2009) Forensic speaker recognition using traditional features comparing automatic and human-in-the-loop formant tracking. *Proceedings of INTERSPEECH (Brighton).* 2343-2346.

De Jong, G., House, J., Cook, N. & Young, A. (2005) The speaker discriminating power of the final fall: spontaneous speech. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Marrakesch*.

De Jong, G., McDougall, K., Hudson, T. & Nolan, F. (2007) The speaker discriminating power of sound undergoing historical change: A formant-based study. *Proceedings of the International Congress of Phonetic Sciences 16 (Saarbrücken)*. 1813-1816. http://www.icphs2007.de/conference/Papers/1542/1542.pdf

De Jong, G., McDougall, K. & Nolan, F. (2007) Sound change and speaker identity: An acoustic study. In Müller (ed.) *Speaker Classification II: Selected Project*s. Berlin: Springer. pp. 130-141.

De Leeuw, E. (2007) Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19: 85-114.

Dellwo, V. & Koreman, J. (2008) How speaker idiosyncratic is measurable speech rhythm? *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics*, Lausanne.

Dellwo, V., Leemann, A. & Kolly, M.J. (2015) Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America* 137(3): 1513-1528. http://dx.doi.org/10.1121/1.4906837.

Duckworth, M. & McDougall, K. (2012) Developing disfluency profiles for individual speakers: A study of Standard Southern British English. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Santander*.

Enzinger, E. (2009) *Formant Trajectories in Forensic Speaker Recognition*. Diplomarbeit, Universität Wien. http://othes.univie.ac.at/8239/1/2009-12-18_0504507.pdf

Eriksson, A. (2010) The disguised voice: Imitating accents or speech styles and impersonating individuals. In: Llamas & Watt (eds.) *Language and Identities*. Edinburgh: Edinburgh University Press. 86-96.

Eriksson, A. (2012) Aural/acoustic vs. automatic methods in forensic phonetic case work. In: Neustein & Patil (eds.) *Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism*. Berlin: Springer. 41-69.

Eriksson, E.J. & Sullivan, K. (2008) An investigation of the effectiveness of a Swedish glide+vowel segment for speaker discrimination. *The International Journal of Speech, Language and the Law* 15: 51-66.

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

Fischer-Weppler, J. & Schiel, F. (2010) The effect of the 'telephone situation' on formant frequencies. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics*, *Trier*.

Flanagan, J.L. (1972) *Speech Analysis, Synthesis, and Perception*. New York: Springer Verlag.

Foulkes, P., Gareth, C. & Hughes, S. (2004) Sociolinguistic and acoustic variability in filled pauses. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Helsinki.*

Foulkes, P. & French, P. (2012) Forensic speaker comparison: A linguistic-acoustic perspective. In: Solan & Tierma (eds.) *Oxford Handbook of Language and Law.* Oxford: OUP. 557-572.

French, P. (1994) An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics* 1: 169-181.

French, P. & Harrison, P. (2007) Position statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases. *The International Journal of Speech, Language and the Law* 14: 137-144.

French, P., Nolan, F., Foulkes, P., Harrison, P. & McDougall, K. (2010) The UK position statement on forensic speaker comparison: a rejoinder to Rose and Morrison. *The International Journal of Speech, Language and the Law* 17: 143-152.

French, P. & Stevens, L. (2012) Forensic Speech Science. In: Anne-Knight & Jones (eds.) *The Bloomsbury Companion to Phonetics*. London: Continuum.

Fridland, V., Kendall, T. & Farrington, C. (2014) Durational and spectral differences in American English vowels: Dialect variation within and across regions. *Journal of the Acoustical Society of America* 136 (1): 341-349. http://dx.doi.org/10.1121/1.4883599

Gfroerer, S. (2003) Auditory-instrumental forensic speaker recognition. *Proceedings of EUROSPEECH (Ghent).* 705-708.

Gfroerer, S. & Wagner, I. (1995) Fundamental frequency in forensic speech samples. *Studies in Forensics* 64: 41-48.

Gold, E. & French, P. (2011) International practices in forensic speaker comparison. *Journal of Speech, Language and the Law* 18: 293-307.

Gold, E. & French, P. (2019) International practices in forensic speaker comparison: second survey. *Journal of Speech, Language and the Law* 26(1): 1-20.

Goldman-Eisler, F. (1968) *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Gordon, M. & Ladefoged, P. (2001) Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29: 383-406.

Greisbach, R., Esser, O. & Weinstock, C. (1995) Speaker identification by formant contours. In Braun & Köster (eds.) *Studies in Forensic Phonetic*s. Trier: Wissenschaftlicher Verlag Trier. 49-55.

Grigoras, C., Jessen, M. & Becker, T. (2009) Forensic speaker verification using long term formant distributions and likelihood ratios. *5th Conference of the European Academy of Forensic Science*.

Grosjean, F. (1980) Temporal variables within and between languages. In Dechert & Raupach (eds.) *Temporal Variables in Speech. Studies in Honor of Frieda Goldman-Eisler*. The Hague: Mouton. 39-53.

Guillemin, B.J. & Watson, C. (2008) Impact of the GSM mobile phone network on the speech signal. *The International Journal of Speech, Language and the Law* 15: 193-218.

Gut, U. (2007) Foreign Accent. In Müller (ed.) *Speaker Classification I: Fundamentals, Features, and Methods*. Berlin: Springer. 75-87.

Gut, U. (2009) Non-native speech. *A Corpus-Based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt/M.: Peter Lang.

Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet (1999) Cambridge: Cambridge University Press.

Harrison, P. (2001) GSM interference cancellation for forensic audio: a report on work in progress. *Forensic Linguistics* 8: 9-23.

Harrison, P. (2014) Making Accurate Formant Measurements: An Empirical Investigation of the Influence of the Measurement Tool, Analysis Settings and Speaker on Formant Measurements. PhD Thesis: University of York.

Haspelmath, M., Dryer, M., Gil, D. & Comrie, B. (eds.) (2005) *The World Atlas of Language Structures*. Oxford: OUP.

Hawkins, S. & Stevens, K.N. (1985) Acoustic and perceptual correlates of the non-nasal-nasal distinctions of vowels. *Journal of the Acoustical Society of America* 77: 1560-1575.

Hazen, K. (2006) Idiolect. In Brown (ed.) *Encyclopaedia of Language and Linguistics*. Oxford: Elsevier [2nd ed.] 5. 512-514.

He, L. & Dellwo, V. (2017) Between-speaker variability in temporal organizations of intensity contours. Journal of the Acoustical Society of America 141 (5): 488-494. http://dx.doi.org/10.1121/1.4983398.

Hess, W. (2008) Pitch and voicing determination of speech with an extension towards music signals. In Benesty, Mohan Sondhi & Huang (eds.) *Springer Handbook of Speech Processing*. Berlin: Springer. 181-211.

Hillenbrand, J., Cleveland, R.A. & Erickson, R.L. (1994) Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37: 769-778.

Hirson, A. (1995) Human Laughter – A forensic phonetic perspective. In Braun & Köster (eds.) *Studies in Forensic Phonetics*, Trier: Wissenschaftlicher Verlag Trier. 77-86.

Hollien, H. (1990) The Acoustics of Crime: The New Science of Forensic Phonetics. New York: Plenum Press.

Hollien, H. (2002) *Forensic Voice Identification*. San Diego: Academic Press.

Hollien, H., Hollien, P.A. & de Jong, G. (1997) Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America* 102: 2984-2991.

Hollien, H., Didla, G., Harnsberger, J.D. & Hollien, K.A. (2016) The case for aural perceptual speaker identification. *Forensic Science International* 269: 8-20.

Hudson, T., McDougall, K. & Hughes, V. (2021) Forensic phonetics. In Setter and Knight (eds.) *Cambridge Handbook of Phonetics.* Cambridge: CUP. 631-656.

Hughes, V., Foulkes, P. & Wood, S. (2016) Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law* 23: 99-132.

Hughes, V., Harrison, P., Foulkes, P., French, P. & Gully, A. (2019) Effects of formant setting and channel mismatch on semi-automatic systems in forensic voice comparison. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 3080-3084.

Hughes, V., Cardoso, A., Foulkes, P., French, P., Harrison, P. & Gully, A. (2019) Forensic voice comparison using long-term acoustic measures of voice quality. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 1455-1459.

Jacewicz, E., Fox, R.A. & Wei, L. (2010) Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America* 128: 839-850.

Jessen, M. (1997) Phonetic manifestations of cognitive and physical stress in trained and untrained police officers. *Forensic Linguistics* 4: 125-147.

Jessen, M. (1997) Speaker-specific information in voice quality parameters. *Forensic Linguistics* 4: 84-103.

Jessen, M. (2007a) Forensic reference data on articulation rate in German. *Science and Justice* 47(2): 50-67.

Jessen, M. (2007b) Speaker classification in Forensic Phonetics and Acoustics. In Müller (ed.) Speaker Classification I: Fundamentals, Features, and Methods. Berlin: Springer. 180-204.

Jessen, M. (2008a) Forensic Phonetics. *Language and Linguistics Compass* 2: 671-711.

Jessen, M, (2008b) Categorical vs. continuous variations between speakers. Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Lausanne.

Jessen, M. (2009) Forensic phonetics and the influence of speaking style on global measures of fundamental frequency. In Grewendorf & Rathert (eds.) *Formal Linguistics and Law*. Berlin: Mouton de Gruyter. 115-139.

Jessen, M. (2010) The forensic phonetician: Forensic speaker identification by experts. In Coulthard & Johnson (eds.) *The Routledge Handbook of Forensic Linguistics*. London: Routledge. 378-394.

Jessen, M. (2018) Forensic voice comparison. In Visconti (ed.), *Handbook of Communication in the Legal Sphere*. Berlin: Mouton de Gruyter. 219-255.

Jessen, M. (2021) Speaker profiling and forensic voice comparison. The auditory-acoustic approach. In Coulthard, May & Sousa-Silva (eds.), *The Routledge Handbook of Forensic Linguistics* [2nd ed.]. London: Routledge. 382-399.

Jessen, M., Köster, O. & Gfroerer, S. (2003) Effect of increased vocal effort on average and range of fundamental frequency in a sample of 100 German-speaking male subjects. *Proceedings of the International Congress of Phonetic Science 15 (Barcelona)*. 1623-1626.

Jessen, M., Köster, O. & Gfroerer, S. (2005) Influence of vocal effort on average and variability of fundamental frequency. *The International Journal of Speech, Language and the Law* 12: 174-213.

Johnson, K. (2012) *Acoustic and Auditory Phonetics* [3rd ed.]. Chichester: Wiley-Blackwell.

Johnstone, B. (1996) The Linguistic Individual. Self-Expression in Language and Linguistics. Oxford: OUP.

Jongman, A., Wayland, R. & Wong, S. (2000) Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108: 1252-1263.

Junqua, J.C. (1996) The influence of acoustics on speech production: a noise-induced stress phenomenon known as Lombard reflex. *Speech Communication* 20: 13-22.

Juste, F.S. & Furquim de Antrade, C.R. (2011) Speech disfluency types of fluent and stuttering individuals: Age effects. *Folia Phoniatrica et Logopaedica* 63: 57-64.

Keating, P.A. (1990) Phonetic representation in a generative grammar. *Journal of Phonetics* 18: 321-334.

Kienast, M. & Glitza, F. (2003) Respiratory sounds as an idiosyncratic feature in speaker recognition. *Proceedings of the International Congress of Phonetic Sciences 15 (Barcelona)*. 1607-1610.

Kim, J. (2019) Individual differences in the production of prosodic boundaries in American English. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 1024-1028.

Kinoshita, Y., Ishihara, S. & Rose, P. (2009) Exploring the discriminatory power of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law* 16: 91-111.

Kirchhübel, C. (2010) The effects of Lombard speech on vowel formant measurements. *The Journal of the Acoustical Society of America* 128: 2395.

Kirchhübel, C., Howard, D.M. & Stedmon, A.W. (2011) Acoustic correlates of speech when under stress: Research, methods and further directions. *The International Journal of Speech, Language and the Law* 18: 75-98.

Klatt, D. & Klatt, L.C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87: 820-857.

Klug, K., Kirchhübel, C., Foulkes, P. & French, P. (2019) Analysing breathy voice in forensic speaker comparison using acoustics to confirm perception. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 795-799.

Koreman, J. (2006) Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119: 582-596.

Köster, O., Schiller, N.O. & Künzel, H.J. (1995) The influence of native-language background on speaker recognition. *Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm)*. 306-309.

Köster, O., Jessen, M., Khairi, F. & Eckert, H. (2007) Auditory-perceptual identification of voice quality by expert and non-expert listeners. *Proceedings of the International Congress of Phonetic Sciences 16 (Saarbrücken)*. 1845-1848.
http://www.icphs2007.de/conference/Papers/1152/1152.pdf

Kraayeveld, J. (1997) Idiosyncrasy in Prosody: Speaker and Speaker Group Identification in Dutch Using Melodic and Temporal Information. PhD Thesis: Catholic University of Nijmegen.

Künzel, H.J. (1989) How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46: 117-125.

Künzel, H.J. (1995) Field procedures in forensic speaker recognition. In Lewis (ed.) *Studies in General and English Phonetics. Essays in Honor of Professor J.D. O'Connor*. London: Routledge. 68-84.

Künzel, H.J. (1997) Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics* 4: 48-83.

Künzel, H.J. (2000) Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics* 7: 149-179.

Künzel, H.J. (2001) Beware of the 'telephone effect': the influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics* 8: 80-99.

Künzel, H.J. (2002) Rejoinder to Francis Nolan's 'The "telephone effect" on formants: a response.' *Forensic Linguistics* 9: 83-86.

Labov, W. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Ladd, D.R. (1996) *Intonational Phonology*. Cambridge: CUP.

Ladefoged, P. (1971) *Preliminaries to Linguistic Phonetics* [2nd ed.]. Chicago: The University of Chicago Press.

Ladefoged, P. (1996) *Elements of Acoustic Phonetics* [2nd ed.]. Chicago: The University of Chicago Press.

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages*. Oxford: Blackwell.

Lane, H. & Tranel, B. (1971) The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research* 14: 677-709.

Laver, J. (1976) Language and non-verbal communication. Reprint in Laver (1991) *The Gift of Speech. Papers in the Analysis of Speech and Voice*. Edinburgh: Edinburgh University Press. 131-146.

Laver, J. (1980) The Phonetic Description of Voice Quality. Cambridge: CUP.

Laver, J. (1994) *Principles of Phonetics*. Cambridge: CUP.

Lawrence, S., Nolan, F. & McDougall, K. (2008) Acoustic and perceptual effects of telephone transmission on vowel quality. *The International Journal of Speech, Language and the Law* 15: 161-192.

Leemann, A., Kolly, M.J. & Dellwo, V. (2014) Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International* 238: 59-67.

Leemann, A. & Kolly, M.J. (2015) Speaker-invariant suprasegmental temporal features in normal and disguised speech. *Speech Communication* 75: 97-122. http://dx.doi.org/10.1016/j.specom.2015.10.002

Liberman, M.Y. & Pierrehumbert, J. (1984) Intonational invariance under changes in pitch range and length. In Aronoff & Oehrle (eds.) *Language Sound Structure*. Cambridge, Mass.: The MIT Press. 157-233.

Lindh, J. (2006) Preliminary descriptive F0-statistics for young male speakers. *Lund University, Dept. of Linguistics & Phonetics Working Papers* 52: 89-92.

Lindh, J. (2017) Forensic Comparison of Voices, Speech and Speakers. Tools and Methods in Forensic Phonetics. PhD Thesis: University of Gothenburg.

Lindh, J. & Eriksson, A. (2007) Robustness of long time measures of fundamental frequency. *Proceedings of INTERSPEECH (Antwerpen)*. 2025-2028.

Linville, S.E. (2001) *Vocal Aging*. San Diego: Singular.

Mackenzie Beck, J. (2005) Perceptual analysis of voice quality: The place of Vocal Profile Analysis. In Hardcastle & Mackenzie Beck (eds.) *A Figure of Speech. A Festschrift for John Laver*. Mahwah: Lawrence Erlbaum Associates. 285-322.

Maddieson, I. (2005) Syllable structure. In Haspelmath, Dryer, Gil & Comrie (eds.) *The World Atlas of Language Structures*. Oxford: OUP. 54-57.

Masthoff, H. (1996) A report on a voice disguise experiment. *Forensic Linguistics* 3: 160-167.

McDougall, K (2004) Speaker-specific formant dynamics: An experiment on Australian English /aɪ/. *The International Journal of Speech, Language and the Law* 11: 103-130.

McDougall, K. (2006) Dynamic features of speech and the characterization of speakers: towards a new approach using formant frequencies. *The International Journal of Speech, Language and the Law* 13: 89-126.

McDougall, K., Nolan, F., Harrison, P. & Kirchhübel, C. (2012) Characterising speakers using formant frequency information: A comparison of vowel formant measurements and long-term formant analysis. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Santander*.

McDougall, K. & Duckworth, M. (2017) Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication* 95: 16-27. http://dx.doi.org/10.1016/j.specom.2017.10.001

McDougall, K. & Duckworth, M. (2018) Individual patterns of disfluency across speaking styles: a forensic phonetic investigation of Standard Southern British English. *The International Journal of Speech, Language and the Law* 25(2): 205-230. http://doi.org/10.1558/ijsll.37241

McDougall, K., Rhodes, R., Duckworth, M., French, P. & Kirchhübel, C. (2019) Application of the 'TOFFA' framework to the analysis of disfluencies in forensic phonetic casework. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 731-735.

Mennen, I., Schaeffler, F. & Docherty, G. (2012) Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America* 131: 2249-2260.

Mielke, J. (2008) The Emergence of Distinctive Features. Oxford: OUP.

Miller, J., Grosjean, F. & Lomanto, C. (1984) Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica* 41: 215-225.

Monsen, R.B. & Engebretson, A.M. (1983) The accuracy of formant measurements: a comparison of spectrographic analysis and linear prediction. *Journal of Speech and Hearing Research* 26(1): 89-97.

Mook, C. & Draxler, C. (2012) A study on the speaker discriminating power of vowels, nasals and fricatives. *Poster presentation at the conference of the International Association for Forensic Phonetics and Acoustics*, Santander.

Moos, A. (2008) Long-Term Formant Distribution (LTF) based on German spontaneous and read speech. *Presentation at the Conference of the International Association for Forensic Phonetics and Acoustics*, Lausanne.

Moos, A. (2010) Long-Term Formant Distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician* 101/102: 7-24.

Moosmüller, S. (1997) Phonological variation in speaker identification. *Forensic Linguistics* 4: 29-47.

Moosmüller, S. (2011) Evaluation of the speech behaviour of reference speakers. *The International Journal of Speech, Language and the Law* 18: 179-205.

Morrison, G.S. (2010) *Forensic Voice Comparison*. In Freckelton & Selby (eds.) *Expert Evidence* (chapter 99). Sydney: Thomson Reuters.

Morrison, G.S., Sahito, F., Jardine, G., Djokic, D., Clavet, S., Berghs, S. & Dorny, C. (2016) INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International* 263: 92-100.

Neuhauser, S. (2008) Voice disguise using a foreign accent: phonetic and linguistic variation. *The International Journal of Speech, Language and the Law* 15: 131-159.

Neuhauser, S. (2011) Variation of glottal activity in French accent imitation produced by native Germans. *The International Journal of Speech, Language and the Law* 18: 207-231.

Nolan, F. (1983) The Phonetic Bases of Speaker Recognition. Cambridge: CUP.

Nolan, F. (1991) Forensic Phonetics. *Journal of Linguistics* 27: 483-493.

Nolan, F. (1994) Auditory and acoustic analysis in speaker recognition. In: Gibbon (ed.) *Language and the Law*. London: Longman. 326-345.

Nolan, F. (1995) Can the definition of each speaker be expected to come from the laboratory in the next decades? *Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm)*. 130-137.

Nolan, F. (1997) Speaker recognition and forensic phonetics. In Hardcastle & Laver (eds.) *The Handbook of Phonetic Sciences.* Oxford: Blackwell. 744-767.

Nolan, F. (2002a) Intonation in speaker identification: an experiment on pitch alignment features. *Forensic Linguistics* 9: 1-21.

Nolan, F. (2002b) The 'telephone effect' on formants: a response. *Forensic Linguistics* 9: 74-82.

Nolan, F. (2005) Forensic speaker identification and the phonetic description of voice quality. In Hardcastle & Mackenzie (eds.) *A Figure of Speech. A Festschrift for John Laver*. Mahwah: Lawrence Erlbaum Associates. 385-411.

Nolan, F. & Grigoras, C. (2005) A case for formant analysis in forensic speaker identification. *The International Journal of Speech, Language and the Law* 12: 143-173.

Nolan, F., McDougall, K., de Jong, G. & Hudson, T. (2009) The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *The International Journal of Speech, Language and the Law* 16: 31-57.

O'Shaughnessy, D. (2008) Formant estimation and tracking. In Benesty, Sondhi & Huang (eds.) *Springer Handbook of Speech Processing*. Berlin: Springer. 213-227.

Ohala, J.J. (1983) The origin of sound patterns in vocal tract constraints. In MacNeilage (ed.) *The Production of Speech*. New York: Springer. 189-216.

Ohala, J.J. (1994) The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, Nichols & Ohala (eds.) *Sound Symbolism*. Cambridge: CUP. 325-347.

Ohala, J.J. & Jaeger, J.J. (eds.) (1986) *Experimental Phonology*. Orlando: Academic Press.

Ostwald, P.F. (1994) Some observations on the function of sound in clinical work. In Hinton, Nichols & Ohala (eds.) *Sound Symbolism*. Cambridge: CUP. 309-324.

Pätzold, M. & Simpson, A.P. (1995) An acoustic analysis of hesitation particles in German. *Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm)*. 512-515.

Park, S.J., Yeung, G., Vesselinova, N., Kreiman, J. & Keating, P.A. (2018) Towards understanding speaker discriminating abilities in humans and machines for text-independent short utterances of different speech styles. *Journal of the Acoustical Society of America* 144(1): 375-386. http://dx.doi.org/10.1121/1.5045323

Quené, H. (2013) Longitudinal trends in speech tempo: The case of Queen Beatrix. *Journal of the Acoustical Society of America* 133(6): 452-457. http://dx.doi.org/10.1121/1.4802892

Redi, L. & Shattuck-Hufnagel, S. (2001) Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29: 407-429.

Reetz, H. & Jongman, A. (2009) Phonetics. Transcription, Production, Acoustics, and Perception. Chichester: Wiley-Blackwell.

Renwick, M.E.L. (2011) Quantifying rhythm: Interspeaker variation in %V. *Journal of the Acoustical Society of America* 130(4): 2567.

Roberts, L. (2011) Acoustic effects of authentic and acted distress on fundamental frequency and vowel quality. *Proceedings of the International Congress of Phonetic Sciences 17 (Hong Kong)*. 1694-1697.

Rose, P. (2002) *Forensic Speaker Identification.* London: Taylor & Francis.

Rose, P. & Morrison, G.S. (2009) A response to the UK Position Statement on forensic speaker comparison. *The International Journal of Speech, Language and the Law* 16: 139-163.

Rostolland, D. (1982a) Acoustic features of shouted voice. *Acustica* 50: 118-125.

Rostolland, D. (1982b) Phonetic Structure of shouted voice. *Acustica* 51: 80-89.

Russ, Ch.V.J. (ed.) (1990) *The Dialects of Modern German*. London: Routledge.

Šalna, B. & Kamarauskas, J. (2010) Evaluation of Effectiveness of Different Methods in Speaker Recognition. *Electronics and Electrical Engineering* 2 (98), 67-70. ISSN 1392-1215. Kaunas Technologija.

Šalna, B. & Bastys, A. (2010) The Use of Group Delay Features of Linear Prediction Model for Speaker. *INFORMATICA* 21 (1), 1-12. ISSN 0868-4952, Institute of Mathematics and Informatics, Vilnius.

San Segundo, E. (2014) Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics. PhD Thesis: Consejo Superior de Investigaciones Científicas, Madrid.

San Segundo, E. & Gómez-Vilda, P. (2015) Evaluating the forensic importance of glottal features through the voice analysis of twins and non-twin siblings. *Language and Law* 1(2): 22-41.

San Segundo, E. & Mompean, J.A. (2017) A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice* 31(5): 644.e11-644.e27. http://dx.doi.org/10.1016/j.jvoice.2017.01.005

San Segundo, E., Foulkes, P., French, J.P., Harrison, P., Hughes, V. & Kavanagh, C. (2019) The use of the vocal profile analysis for speaker characterisation: a methodological proposoal. *Journal of the IPA* 49(3): 353-380.

Schiller, N.O., Köster, O. & Duckworth, M. (1997) The effect of removing linguistic information upon identifying speakers of a foreign language. *Forensic Linguistics* 4: 1-17.

Schötz, S. (2006) *Perception, Analysis and Synthesis of Speaker Age*. Dissertation: University of Lund [ed. as Traveaux de l'Institut de Linguistique de Lund, Vol. 47].

Sharma, N.K., Ganesh, S., Ganapathy, S. & Holt, L.L. (2019) Talker change detection: A comparison of human and machine performance. *Journal of the Acoustical Society of America* 145(1): 131-142. http://dx.doi.org/10.1121/1.5084044

Shriberg, E. (2007) Higher-level features in speaker recognition. In Müller (ed.) *Speaker Classification I: Fundamentals, Features and Methods*. Berlin: Springer. pp. 241-259.

Simpson, A.P. (2007) Phonetic details of nonpulmonic stop release in German: inter- and intraindividual variation. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Plymouth*.

Simpson, A.P. (2009) Phonetic differences between male and female speech. *Language and Linguistics Compass* 3: 621-640.

Simpson, A.P. (2011) Is there a relationship between acoustic vowel space size and fundamental frequency? *Proceedings of the International Congress of Phonetic Sciences 17 (Hong Kong)*. pp. 1854-1857.

Simpson, A.P. & Ericsdotter, C. (2007) Sex-specific differences in f0 and vowel space. *Proceedings of the International Congress of Phonetic Sciences 16 (Saarbrücken)*. pp. 933-936. http://www.icphs2007.de/conference/Papers/1333/1333.pdf

Slifka, J. (2006) Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20: 171-186.

Smirnova, N., Starshinov, A., Oparin, I. & Goloshchapova, T. (2007) Speaker identification using selective comparison of pitch contour parameters. *Proceedings of the International Congress of Phonetic Sciences 16 (Saarbrücken)*. 1801-1804. http://www.icphs2007.de/conference/Papers/1138/1138.pdf

Smorenburg, B.J.L. & Heeren, W. (2019) Speaker variation in Dutch /x/ and /s/ as a function of syllabic position and lip-rounding. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 612-616.

Solé, M.-J., Beddor, P.S. & Ohala, M. (eds.) (2007) *Experimental Approaches to Phonology*. Oxford: OUP.

Sorensen, M.H. (2011) Acoustic and Perceptual Aspects of Speaker-Specific Differences in Speech and their Forensic Implications. PhD Thesis: University Aarhus.

Stevens, K.N. (1971) Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings of the International Congress of Phonetic Sciences 7 (Montreal)*. 206-232.

Stevens, K.N. (1998) *Acoustic Phonetics*. Cambridge, Mass.: The MIT Press.

Stevens, L. & French, P. (2012) Voice quality in Standard Southern British English: distribution of features, inter-speaker variability and the effect of telephone transmission. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Santander*.

Stove, L.M. & Golob, E.J. (2013) Evidence that the Lombard effect is frequency-specific in humans. *Journal of the Acoustical Society of America* 134(1): 640-647. http://dx.doi.org/10.1121/1.4807645

Stuart-Smith, J. & Timmins, C. (2010) The role of the individual in language variation and change. In Llamas & Watt (eds.) *Language and Identities*. Edinburgh: Edinburgh University Press. 39-54.

Tilsen, S. & Arvaniti, A. (2013) Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America* 134(1): 628-639. http://dx.doi.org/10.1121/1.4807565

Titze, I. (1994) *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.

Tsao, Y.-C. & Weismer, G. (1997) Interspeaker variation in habitual speaking rate: evidence for a neuromuscular component. *Journal of Speech, Language, and Hearing Research* 40: 858-866.

Tsao, Y.-C., Weismer, G. & Iqbal, K. (2006a) The effect of intertalker speech rate variation on acoustic vowel space. *Journal of the Acoustical Society of America* 119: 1074-1082.

Tsao, Y.-C., Weismer, G. & Iqbal, K. (2006b) Interspeaker variation in habitual speaking rate: additional evidence. *Journal of Speech, Language, and Hearing Research* 49: 1156-1164.

Tschäpe, N., Trouvain, J., Bauer, D. & Jessen, M. (2005) Idiosyncratic patterns of filled pauses. Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Marrakesch.

Wagner, I. (1995) A new jitter algorithm to quantify hoarseness: an exploratory study. *Forensic Linguistics* 1(1): 18-27.

Wagner, I. (1998) Recent trends in voice disguise. Presentation at the conference of the International Association for Forensic Phonetics, Voorburg.

Wagner, I. (2017) The BKA Standard Operation Procedure of Forensic Speaker Comparison and Examples of Casework. *Presentation at the conference of the International Association for Forensic Phonetics and Acoustics, Split*.

Wagner, I. (2019) Examples of Casework in Forensic Speaker Comparison. *Proceedings of the International Congress of Phonetic Sciences 19 (Melbourne)*. 721-725.

Wagner, I. & Köster, O. (1999) Perceptual recognition of familiar voices using falsetto as a type of voice disguise. *Proceedings of the International Congress of Phonetic Sciences* 14. 1381-1384.

Watt, D. (2010) The identification of the individual through speech. In Llamas & Watt (ed.) Language and Identities. Edinburgh: Edinburgh University Press. 76-85.

Wendahl, R.W. (1966) Some parameters of auditory roughness. *Folia Phoniatrica et Logopaedica* 18: 26-32.

Wenndt, S. (2016) Human recognition of familiar voices. *Journal of the Acoustical Society of America* 140(2): 1172-1183. http://dx.doi.org/10.1121/1.4958682

Wester, M. (2012) Talker discrimination across languages. *Speech Communication* 54: 781-790.

Williams, D. & Escudero, P. (2014) A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *Journal of the Acoustical Society of Americ*a 136(5): 2751-2761. http://dx.doi.org/10.1121/1.4896471

Whitmore, J. & Fisher, S. (1996) Speech during sustained operations. *Speech Communication* 20: 55-7.