Best Practice Manual for the Methodology of Forensic Speaker Comparison

ENFSI-BPM-FSC-01

Version 01 – November 2021

ENFSI´s position on Best Practice Manuals

ENFSI wishes to promote the improvement of mutual trust by encouraging forensic harmonization 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).

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 website of ENFSI (www.enfsi.eu) for update information.

Acknowledgements
Isolde Wagner (Federal Criminal Police Office of Germany, Forensic Science Institute, Germany), Dr. Dagmar Boss (Bavarian State Criminal Police Office, Forensic Science Institute, Germany), Timur Svirava (North-Western Forensic Science Center of the Ministry of Justice, St. Petersburg, Russia), Ivan Siparov (North-Western Forensic Science Center of the Ministry of Justice, St. Petersburg, Russia), Dr. Vincent Hughes (Department of Language and Linguistic Sciences, University of York, UK), Dr. Miriam Rolfes (Berlin State Criminal Police Office, Forensic Science Institute, Germany)

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

The content of this Best Practice Manual for the Methodology of Forensic Speaker Comparison represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Best Practice Manual
for the Methodology of Forensic Speaker Comparison

CONTENT

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. It 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 evidence. Literature with more detailed information of different aspects of the methodology of Forensic Speaker Comparison is provided in §15 ('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. In this series of ENFSI Best Practice Manuals the term BPM has been maintained for reasons of continuity and recognition.

2. SCOPE

The methodology described in this Best Practice Manual is based on a combined procedure of phonetic-linguistic auditory and acoustic analyses of different speech features in Forensic Speaker Comparison. 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 not.

This BPM does not address automatic and semiautomatic speaker recognition. However, reference is made to the ENFSI document 'Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition' [01].

It is important to acknowledge that different methods have advantages and disadvantages. It is possible to combine automatic, semi-automatic and linguistic-phonetic-acoustic methods. The different methods are likely to provide complementary information. However, the specific choice of method or methods is dependent on the suitability of the material (in terms of quality, quantity etc.) as well as on the availability of methods within a given lab and the rules of the jurisdiction in which the expert works.

3. DEFINITIONS AND TERMS

Not required. Explanations are included in the text.

4. RESOURCES

4.1 Personnel

The methodology of Forensic Speaker Comparison 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 particular cases and has the key role for directing and performing the examination. The examination includes auditory and acoustic analyses of submitted speech material, interpreting findings, providing the evidence, writing the expert report and presentation at court.

4.1.1 Practitioner Qualification

A master's degree 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 standards of competence, a documented training program, and processes for assessing whether trainees have achieved the required level of competence.

Competence may be defined as the standards that should be achieved in order for the individual to undertake casework. Personnel will achieve the required level of competence through initial training before being authorised to undertake casework. They should also be subject to regular assessment to ensure that these competencies are maintained and developed.

The assessment of competence can be accomplished through a number of different mechanisms including formal tests, undeclared trials, and peer review of case analysis and reports.

All experts should collect and maintain evidence supporting their ongoing competence and proficiency. In addition to any formal assessment by their organisations, they should read professional literature containing pertinent information relating to forensic phonetics and speaker comparison, take part in appropriate events, e.g. workshops, seminars, training courses, etc., and when possible, actively participate in research and development projects.

4.2 Equipment

The equipment regularly used in Forensic Speaker Comparison are computer devices with specific software for acoustic signal processing with various acoustic devices such as different kinds of loudspeakers, earphones, microphones, sound cards, amplifiers, etc.

In cases involving analogue recordings, appropriate ADC equipment must be used to digitise them.

The equipment inventory must be documented according to the requirements of the laboratory and it must be shown to operate properly before being used in casework.

4.3 Reference Materials

Not applicable

4.4 Accommodation and Environmental Conditions

Forensic laboratories performing examinations in Forensic Speaker Comparison should be designed for efficient and effective operation.

As auditory and acoustic analyses are the important tasks of the methodology, soundproofing is required. Environmental noises should not interfere with auditory and acoustic examinations within the laboratory. Rooms should be furnished with materials for controlled noise level from outside and from ventilation. A noise level lower than 30 dB(A) is strongly recommended.

To avoid further disturbances there should be only one practitioner per room.

4.5 Material and Reagents

Not applicable

5. METHODS

5.1 Principles

The methodology of Forensic Speaker Comparison was developed for the purpose of analysing unknown speakers in telephone interceptions and other forensic and comparing them either among themselves and/or with known speakers.

If there is more than one unknown speaker within one case scenario, unknown speakers must not be pooled automatically. They have to be analysed separately or at least checked in detail before being pooled.

A fundamental principle of this methodology is that a range of speech features can be analysed to capture the many dimensions on which speakers can be distinguished. According to the literature (e.g. [09], [10], [16]), speech features should be as independent from one another 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 are 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 experts play a substantial role. The expert has to make decisions at every step of the speaker comparison analysis using available scientific background information, as well as their 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 and calculations of the speech signal on the other hand. The analyses cover a variety of discriminatory speech features which can be categorised into segmental and supra-segmental features or can be put in parent categories like ‘language’, ‘voice’ or ‘manner of speaking’.

5.2 Speech Feature Analysis

A wide range of linguistic and phonetic features could in principle be analysed in forensic speaker comparison. Many can be analysed both auditorily and acoustically. Those chosen for analysis will 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 their susceptibility to within-speaker variability (due to e.g. emotion, illness, drugs, alcohol). Below some of the most commonly analysed features and those with the greatest diagnostic value are described:

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.

Such analyses help answer the questions about which language was used, and whether it was a native or a second language. Furthermore, the analyses can give information about the type and degree of dialectal or regional influences, and about the possible presence of a foreign accent or a certain type of slang. Socially or individually distinctive features like specific wording or pronunciation can also be provided.

The analyses are conducted on both a segmental and supra-segmental basis. There is investigation into the articulation of vowels and consonants on the one hand, and on prosodic patterns like intonation on the other hand. Additionally, there are linguistic analyses of lexical and grammatical realisations in speech. The whole analysis process is auditory phonetic-linguistic oriented with acoustic analysis conducted where necessary and possible. This results in a detailed description of individual features against common norms and standards in the field of language and speech.

If the expert is a non-native speaker of the language under analysis, he or she has to co-operate with a native speaker or a person with thorough knowledge of the language (e.g. an interpreter).

5.2.2 Fundamental Frequency and Variation

Fundamental frequency (F0) refers to the approximate frequency of the (quasi-)periodic structure of voiced speech signals.

The human voice is not a pure tone but is composed of an F0 and a series of higher frequencies called upper harmonics.

F0 is nearly always reported as an average measure, usually expressed in Hz, but in some cases the average duration of a period is reported instead. The frequency in Hz is thus reciprocal of the period (1 divided by the period in seconds).

Typical ranges of F0 are 80-200 Hz for men and 150-400 Hz for women who normally have higher F0 due to shorter vocal fold length. The mean values are reported to change slightly with age. F0 range for an adult is reported to be variable and comprises 1.5-2 octaves. Intra-speaker variability is reported to be affected by the conditions under which the speech is being produced (noise condition, channel, emotions etc.).
Variables that influence F0 include tension of the cord, force of glottal closure indicated by the glottal resistance, and expiratory air pressure.

The characteristics of the F0 might be analysed in a Forensic Speaker Comparison case both auditorily (through pitch level) and instrumentally (there is a number of F0 detection techniques e.g. based on the cepstrum of the speech signal). The following parameters might be analyzed: mean, median, max and min values, standard deviation and so on.

5.2.3 Voice Quality

The term voice quality implies quasi-permanent characteristics that are typical for the overall sound of a speaker's voice. These characteristics can be 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 typical for a speaker's voice both in terms of the extent of oscillations and the frequency of oscillations. Categories for the description of phonation are, for example, a harsh, tense, breathy or creaky voice.
The source signal produced by phonation is then filtered by the vocal tract (made up of the pharyngeal, oral and nasal cavities). The anatomic-morphologic conditions of the vocal tract and the individual neuro-muscle stereotypes determine the resonance characteristics that are typical for each individual. Categories for the description of vocal tract influence are, for example, nasality, lip rounding or spreading, and close or open jaw.

The characteristics of voice quality are analysed both auditorily and by the evaluation of spectrograms or acoustic measurements. In doing so, the individual characteristics can be described in terms of frequency of appearance, degree and representativity. The reference is the so-called modal voice of Laver [12] and analysis is conducted in terms of deviations from this modal 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. They play a very important role in speech signal analysis as their positions are defined by the geometry and specific shape of the individual vocal tract of the speaker. The relative position of articulators is reflected in the shape of the filter that is applied to the speech source signal. The source of the speech wave can be the vocal folds (vocalised sounds) or turbulent noise of different localisations (whisper, fricatives, spirants, etc.) [03]. Vocalised sounds are mostly used for formant analysis as they provide rich signal source spectrum, so its filtering by the vocal tract can be finely recognised over the spectrum. In this case, the full set of fundamental tone sub-harmonics fill the signal spectrum and is filtered by the frequency response of the vocal tract.

Different vocal tract configurations (different articulations according to different mutual positions of articulation organs: tongue, jaws, teeth, lips, etc.) provide different spectral maxima positions, which are responsible for maintaining phonetic contrast between different phonemes.

There is a certain correlation between formants' position in the speech signal spectrum and phonemes. Thus, the speech signal spectrum is determined both by the phonetics of the language or variety on the one hand and individual shape of the speakers' vocal tract on the other. All speakers have to adapt their pronunciation patterns to the spoken language environment and train their neuro-muscle skills in such a way that their sounds (in general, lower formant positions F1, F2, and in some cases F3) match the language- or variety-specific phonetic patterns and are recognised by the others.

Two general approaches to statistical analysis of formants' distribution are normally used:

1) Single vowel formant analysis is the method that is focused on the collection of formant statistics for a certain vowel or articulation and their comparison for the speakers under investigation. When comparing the vowels with the same lower formant positions, more speaker-specific information is in general contained within the higher formants.

2) Long term formant distribution (LTFD) is the method used to capture the distribution of values for each formant of a speaker over a given speech recording. For a given formant (F1, F2, F3 etc.) measurements for all vowels produced by a single speaker are taken across the entire recording or relevant portions of the recording. This means that each formant produced by a speaker can be reduce to a mean and a standard deviation. 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.

Channel impact: As the speech signal is transmitted through different analogue channels and digital processing during the recording the impact of these transformations should be considered during the formant analysis. Microphone frequency response, telephone line and radio channel are examples of additional analogue filtering that can affect formant values. GSM

codecs are an example of digital processing that also affects the formants positions as it stores 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 analysis and investigation.

Precision: The precision of formant measurement in general is pre-defined with the F0 value being the sub-harmonic signal source to be filtered by vocal tract frequency response. The main approaches to measuring formants positions are: wideband spectrogram, cepstrally smoothed spectrogram and LPC-spectrogram. The compromise between spectral resolution, formant peak factor and noise robustness of the representation determines which choice produces the best result.

Formant positions statistics should be treated considering their representativeness, providing well-recognised patterns and repeatable results.

5.2.5 Articulation Rate

Tempo is the speed (or rate) at which an individual produces speech. It is generally analysed as either speaking rate or articulation rate. Speaking rate represents 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 (Laver [13]).

While it is possible to judge qualitatively (using subjective labels like 'fast' or 'slow'), speech tempo is commonly analysed quantitively; most commonly measured as the number of syllables per second. When measuring tempo, it is necessary for the analyst to 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).

A number of studies have produced a variety of reference values for speech tempo. Goldman-Eisler's [04] study is still widely cited and reports articulation rates of between 4.4 to 5.9 syllables per second being based on eight speakers in spontaneous conversations. Average articulation rates of 5.9 syllables per second are reported by Künzel [11] for German. It is well known that tempo is affected by factors such as speaking style and so displays considerable within-speaker variability. Speech tempo is therefore only likely to be of value where speakers are at the extreme ends of the population distribution.

5.2.6 Hesitation Phenomena and other non-pathological speech disfluencies

Hesitation phenomena are those features of speech, voice and manner of speaking which either occur unintentionally or are used deliberately when a speaker hesitates/pauses before or within an utterance.

When analysing and describing hesitation phenomena, it should, in general, be stated how often they appear, how long they are and whether they are accompanied by a particular voice quality or by features like a glottal stop. Common hesitation phenomena are, for example:

(1) Use of hesitation particles, such as *uh* / *uhm* in English or *äh* / *ähm* in German.
(2) Lengthening of vowels and consonants. Here it should be stated which vowels and/or consonants are affected most often.
(3) Occurrence of silent pauses. It should also be checked whether their use is related to the content of the utterance (emphasising the following part of the utterance, for example) or if it is erratic.
(4) 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 or produce a completely new utterance, or repeat parts of an utterance (sentences, single words, syllables and/or sounds - the last two sometimes perceived as light stuttering).

Most speakers tend to use a certain inventory of hesitation phenomena. The frequency of occurrence and combination of these phenomena can be of speaker-specific value 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 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 classed as a disfluency of which there are many kinds (see McDougall and Duckworth [14] for more discussion on the analysis of disfluencies for non-pathological speakers in the forensic context). However, pathology can also affect the phonetics of speech production, especially where an individual has some physical impairment (such as cleft palate or laryngitis) 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 supralaryngeal (such as fronting of coronal sibilants due to a lisp) levels, are most likely to be affected. Indeed, the vocal profile analysis (VPA see §5.2.3; Laver [12]) scheme, which is now widely used by forensic phoneticians as a tool in speaker comparison cases, was originally developed to document pathological voice qualities and vocal settings (where ratings of 4-6 indicate pathology).

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.

5.3 Peer Review

The findings of a speaker comparison analysis and the final report should be checked by a second expert if it is possible and admitted by local legislation. This peer review should be documented in the case file.

6. VALIDATION AND ESTIMATION OF UNCERTAINTY OF MEASUREMENT

6.1 Validation

The laboratory should demonstrate, where possible, that the methods used in Forensic Speaker Comparison cases are validated. Essential to validation is method development. This involves:

- Selection of an appropriate pool of features for analysis (i.e. the features that could in principle be useful; as outlined in §5)
- Selection of specific features for the given case (dependent not only on diagnostic value, but also what is available for analysis within the samples themselves)
- Knowledge of what software tools are appropriate and reliable for different types of analysis features.

Method development for Forensic Speaker Comparison is 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 practitioners (e.g. structured training programmes, conferences).

Validation itself involves being able to demonstrate that the method applied is able to separate same-speaker and different-speaker pairs with a given level of accuracy. This can be demonstrated through proficiency tests (PTs) and collaborative exercises (CEs).

As part of the validation process, it is essential to establish whether the method meets the required accuracy, precision, repeatability, reproducibility and robustness of its intended use. The results of the validation process should be produced in a report as per Standard Operating Procedures (SOPs). Validation should be as extensive as is necessary for the application or field. New methods and tools should also be validated as they become available. Over time, it is necessary to re-validate methods. Where the techniques or procedures have been validated elsewhere, the laboratory is required to carry out a verification exercise to demonstrate that it can achieve the same quality of results in its own environment. 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' [08].

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 Forensic Speaker Comparison. The 'Guidance on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI' [07] 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 the methodology of Forensic Speaker Comparison. The test material is speech, which is language dependent, and thus varies through different countries and regions. Proper 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 cannot be organised as a joint venture for the ENFSI EWG of Forensic Speech and Audio Analysis. These PTs and CEs have to be organised at a national or regional level. The organisation of such PTs and CEs and the frequency of participation follows the rules of different ENFSI member laboratories and their accreditation bodies.

7.2 Quality Controls

Within the methodology of Forensic Speaker Comparison, the procedure of feature analyses, evaluation of results, and reporting should be documented with regard to the policy of local practices and whenever possible in SOPs. The SOPs should be included in a training programme, regularly reviewed, and updated if necessary.

When using the methodology of Forensic Speaker Comparison for casework it is important to have appropriate 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.) to provide transparency and traceability.

Since the methodology of Forensic Speaker Comparison is a human-based method (dependent on the competence and knowledge of the expert) it is recommended that its application to casework is carefully checked by a second expert as part of peer review (also see §5.3).

Management reviewing, in the sense of plausibility check, follows the rules of the EWG member laboratories.

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 transmission of the material from the customer to the expert and back has to be documented thoroughly (referred to as the chain of custody).

In most cases the material comes on a digital base either via data transfer or on a digital carrier. The original material should be stored as a 1 to 1 copy before any processing of the data (e.g. channel separation, conversion, editing etc.).

In rare cases the material can be handed in on an analogue medium, like a wheel tape or an audio cassette. In such cases, the audio signal has to be digitised in an appropriate way. A suitable device has to be chosen and checked with a test tape to assess whether it functions correctly. Then the material has to be digitised with a sampling rate of minimum 44.1 kHz, 16 bit, format: PCM Wave.

When working with cassettes it has to be made sure that the safety tabs (writing protection) are removed to prevent the content from being erased unintentionally. If the safety tabs have to be removed by the practitioner, they should be kept with the cassette as they belong to the evidence.
Each step has to be documented.
For the speech feature analyses themselves, only the digital copy will be used.
To check whether the original has been recorded at the right speed (in old tapes often deviations from the standard speed occur so that the speech signal is falsified in this respect) an ENF component which could have been produced by the recording device can be looked for and it has to be checked whether the ENF component is at its expected standard (50/60 Hz or harmonics). If necessary, it has to be adjusted to prevent false findings.

In all cases the material submitted to the laboratory has to be registered, checked for correctness, completeness, and prioritisation (if needed) according to the laboratory's regulations and the customer's request. Confidentiality of the case material must also be considered. If necessary, the customer must be consulted. Possible questions on the material have to be clarified.

9. INITIAL ASSESSMENT

In general, all casework should undergo an initial case assessment to determine the suitability of the material for examination and the applicability of methods before any examination is undertaken.

9.1 Suitability/Applicability

Every speech feature analysis within the methodology of FSC has its own scope and range of efficient application. The suitability of the speech material submitted for examination determines the applicability of the methods used. General criteria of speech materials' suitability for the investigation include speech duration, speech representativity and acoustic quality, e.g. speech signal to background signal ratio (energy and spectrum).

Suitability can be established during preliminary analysis of the speech material or, for some of the methods, during the analysis itself. Speech suitability is evaluated to establish that the methods applied to the material provide the customer with a reliable result. A lack of suitability of speech material can lead to particular feature being excluded from the analysis or a full rejection of the examination.

9.2 Comparability of Speech Material

The comparability of speech at the questioned audio and the reference material plays an important role in Forensic Speaker Comparison.

The circumstances, in which the material was recorded, should be assessed both from the technical point of view and from the point of view of situational factors determined by the whole communication process. Evaluating the degree of matches and mismatches allows the expert to estimate the influence on the features extracted.

Technical parameters to be checked for comparability are the recording format (e.g. compression and coding of the signal that might affect audio quality during recording), the recording channel (e.g. telephone, microphone, radio), and environmental conditions given

from the recording process (e.g. the level and composition of background noises or channel distortions affecting the recorded signal or conditioning any particular features of the communication).

The communication situation has to be checked for comparability in many respects. There can be monologues or dialogues, conversation with or without visual contact, etc. A given situation can influence the emotions of a speaker having significant impact on linguistic and phonetic features, such as e.g. the dynamics of the voice, lexical choices, prosodic patterns or cause a general physiological tension that could determine the speech behavior.

Furthermore, mismatch of spoken languages between the questioned and the reference material could play a substantial role. It limits the number of phonetic-linguistic features that could, in principle, be analysed. Caution should therefore be exercised in analysing cases involving language mismatch.

For each feature analysed there should be an assessment of the comparability of the questioned and reference material. Limited comparability between questioned and reference material can limit the analyses or even lead to a complete rejection of the whole Forensic Speaker Comparison examination.

If there is no reference material submitted or if the reference material is not sufficient it can be recorded by the expert. In this case the recommendations in Appendix 1 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. EVALUATION AND INTERPRETATION

After detailed analyses of various discriminatory speech features (the most commonly are described in §5), all similarities and differences observed between the compared speakers are evaluated by the expert. The evaluation process must include considerations about the degree of suitability and comparability of the material at the same time as considerations about the characteristics of features and their frequency of occurrence in connection with inter- and intraindividual speaker variations. The interpretation is carried out both by evaluating results for each individual feature and, on more holistic level, by considering the outcome of the whole examination.

Conclusions in Forensic Speaker Comparison cases are in many respects subjective and based to a large extent upon the competence (education, training and experience) of the expert. This is due to the challenges of empirical and statistical analysis for many speech features (e.g. in the context of language, dialect etc.) and, often, due to issues of channel and technical mismatch between samples.

At present, there is no universally accepted scaling for reporting conclusions in forensic speaker comparison cases. The type and range of scales differ widely between different laboratories (see e.g. [05], [06], [15] in the literature). Irrespective of the conclusion scale used, the whole Forensic Speaker Comparison examination including evaluation and interpretation should undergo validation and quality assurance processes. Statements should be considered and expressed with thorough and care.

A general framework proposed for evaluation and interpretation is described in the 'ENFSI Guideline for Evaluative Reporting in Forensic Science' [02].

13. PRESENTATION OF EVIDENCE

Evidence can be presented to the court either orally or in writing. Presentation of evidence should clearly state the results of any evaluation and interpretation of the examination.
Written reports should include all the relevant information in a clear, concise, structured and unambiguous manner as required by the relevant jurisdiction. This is also done according to the relevant laboratory regulations. Written reports must be peer reviewed.

14. HEALTH AND SAFETY

During the process of auditory analysis, it must be taken into account that both, single loud events of more than 120 dB(A) and permanent loudness of more than 80 dB(A) can cause psychological stress, tinnitus, hearing impairment or hearing loss.
Experts listening over an extended time span can face similar risks for health. For this reason, pausing is required depending on the equipment that is used. Experience has shown that, when listening to loud speakers, there should be a minimum of 30 minutes brake after at most two hours of intense auditory analysis. When using headphones there should be more frequent, but much shorter sequences of intense auditory analysis of at most one hour. Headphones are often used with a louder volume. Extended use of headphones influences the general perception in spatial hearing.

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 stress, anxiety, and traumas. In such situations psychological support should be taken.

15. BIBLIOGRAPHY

Referred literature and documents (valid in their latest version):

[01] 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

[15] European Network of Forensic Science Institutes (2015): ENFSI guideline for evaluative reporting in forensic science: Strengthening the Evaluation of Forensic Results across Europe. https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf

[16] Fant, G. (1960): *Acoustic theory of speech production*. The Hague: Mouton.

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

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

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

[04] Guidance on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI; QCC-PT-001;
http://enfsi.eu/wp-content/uploads/2017/07/QCC-PT-001-_-Guidance-on-PT-CE.pdf
and
http://enfsi.eu/wp-content/uploads/2017/06/External-proficiency-tests-and-collaborative-exercises-PT-CE-providers-28Nov2018.xlsx
and
_exercises-PT-CE-provided-by-ENFSI-01Oct2018.xlsx

[03] Guidelines for the single laboratory Validation of Instrumental and Human Based Methods in Forensic Science; QCC-VAL-002;
http://enfsi.eu/wp-content/uploads/2017/06/Guidance-QCC-VAL-002.pdf
and GUIDANCE – ANNEX;
http://enfsi.eu/wp-content/uploads/2017/06/Guidance-QCC-VAL-002-Annex.pdf

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

[06] Jessen, M. (2021): Speaker profiling and forensic voice comparison. The auditory-acoustic approach. In: The Routledge Handbook of Forensic Linguistics, 382-399.

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

[08] Laver, J. (1980): *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

[09] Laver, J. (1994): *Principles of Phonetics*. Cambridge: Cambridge University Press.

[10] McDougall, K. & Duckworth, M. (2017): Profiling fluency: an analysis of individual variation in disfluencies in adult males. *Speech Communication* 95, 16-27.

[12] Morrison, G, 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.

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

There are many books, journals and individual papers published on the subject of Forensic Speaker Comparison. The following list contains a collection of further publications that relate to the methodology of Forensic Speaker Comparison (forensic phonetic-linguistic auditory and acoustic approach).

[x] 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.

[x] Bahr, R. Huntley & 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.

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

[x] Batliner, A., Kießling, A., Burger, S. & Nöth, E. (1995): Filled pauses in spontaneous speech. *Proceedings of the International Congress of Phonetic Sciences* 13 (Stockholm), Vol. 3, 472-475.

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

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

[x] 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

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

[x] 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. Proceeding of the International Congress of Phonetic Sciences 17 (Hong Kong), 356-359. http://www.icphs2011.hk/resources/OnlineProceedings/RegularSession/Bradlow/Bradlow.pdf

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

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

[x] 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

[x] 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.

[x] 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.

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

[x] 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.

[x] 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.

[x] Cao, H. & Wang, Y. (2001): A forensic aspect of articulation rate variation in Chinese. Proceeding of the International Congress of Phonetic Sciences 17 (Hong Kong), 396-399. http://www.icphs2011.hk/resources/OnlineProceedings/RegularSession/Cao,%20Honglin/Cao,%20Honglin.pdf

[x] 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

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

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

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

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

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

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

[x] 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.

[x] 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* 2009, Brighton, 2343-2346.

[x] 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.

[x] De Jong, G., McDougall, K., Hudson, T. & Nolan, F. (2007a): 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

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

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

[x] 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.

[x] 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

[x] 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.

[x] Enzinger, E. (2010): Formant Trajectories in Forensic Speaker Recognition. Diplomarbeit, Universität Wien.
http://othes.univie.ac.at/8239/1/2009-12-18_0504507.pdf

[x] 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.

[x] 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.

[x] 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.

[x] 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.

[x] Flanagan, J.L. (1972): *Speech analysis, synthesis, and perception*. New York: Springer Verlag.

[x] 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.

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

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

[x] 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.

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

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

[x] 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

[x] Gfroerer, S. (2003): Auditory-instrumental forensic speaker recognition. *Proceedings of EUROSPEECH* 2003, Genf, 705-708.

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

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

[x] 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.

[x] 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.

[x] 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.

[x] 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.

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

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

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

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

[x] 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 Dissertation. University of York.

[x] Haspelmath, M., Dryer, M. Gil, D. & Comrie, B. (eds.) (2005): *The World Atlas of Language Structures*. Oxford: Oxford University Press.

[x] 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.

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

[x] 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

[x] 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.

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

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

[x] Hollien: H (1990): *The Acoustics of Crime: The New Science of Forensic Phonetics*. New York: Plenum Press.

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

[x] 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.

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

[x] Hudson, T., McDougall, K. & Hughes, V. (to appear, 2021) Forensic speech science. In Setter, J. and Knight, R-A. (eds.) *Cambridge Handbook of Phonetics.* Cambridge: CUP.

[x] Hughes, V., Foulkes, P. and 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.

[x] 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.

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

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

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

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

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

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

[x] 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.

[x] 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.

[x] 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.

[x] 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.

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

[x] Johnstone, B. (1996): *The Linguistic Individual. Self-Expression in Language and Linguistics*. Oxford: Oxford University Press.

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

[x] 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.

[x] 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.

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

[x] 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.

[x] 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.

[x] 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.

[x] Kirchhübel, Ch. (2009): The effects of Lombard Speech on Vowel Formant Measurements. MSc-Thesis, University of New York.

[x] Kirchhübel, Ch., 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.

[x] 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.

[x] Klug, K., Kirchhübel, Ch., Foulkes, P. & French, P. (2919): Analysing breathy voice in forensic speaker comparison using acoustics to confirm perception. *Proceedings of the International Congress of Phonetic Sciences* 19 (Melbourne), 795-799.

[x] 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.

[x] 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), Vol. 3, 306-309.

[x] 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

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

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

[x] 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.

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

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

[x] 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.

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

[x] Labove, W. (1972): *Sociolinguistic Patterns*. Philadelphia: University of Pennsilvania Press.

[x] Ladd, D.R. (1996): *Intonational Phonology*. Cambridge: Cambridge University Press.

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

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

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

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

[x] 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.

[x] Laver, J. (1994): *Principles of Phonetics*. Cambridge: Cambridge University Press.

[x] 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.

[x] 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.

[x] 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

[x] 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.

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

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

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

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

[x] 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.

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

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

[x] 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.

[x] 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.

[x] McDougall, K., Nolan, F., Harrison, P. & Kirchhübel, Ch. (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.

[x] 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

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

[x] McDougall, K., Rhodes, R., Duckworth, M., French, P. & Kirchhübel, Ch. (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.

[x] 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.

[x] Mielke, J. (2008): *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

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

[x] 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.

[x] Mook, C. & Draxler, Ch. (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.

[x] 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.

[x] Moos, A. (2010): Long-Term Formant Distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician* 101/102, 7-24. http://www.isphs.org/Phonetician/Phonetician_101.pdf

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

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

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

[x] 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.

[x] 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.

[x] Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

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

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

[x] 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), Vol. 3, 130-137.

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

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

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

[x] 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.

[x] 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.

[x] 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.

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

[x] 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.

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

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

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

[x] Pätzold, M. & Simpson, A. (1995): An acoustic analysis of hesitation particles in German. *Proceedings of the International Congress of Phonetic Sciences* 13 (Stockholm), Vol. 3, 512-515.

[x] 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

[x] 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

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

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

[x] Renwick, M.E.L. (2011): Quantifying rhythm: Interspeaker variation in %V. *Journal of the Acoustical Society of America* 130, No. 4, Pt. 2, 2567 [Abstract].

[x] 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.

[x] 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.

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

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

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

[x] 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.

[x] 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.

[x] 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. Ph.D. Dissertation, Consejo Superior de Investigaciones Científicas.

[x] 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.

[x] 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

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

[x] 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.

[x] 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].

[x] 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

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

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

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

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

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

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

[x] 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

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

[x] Solé, M.-J., Beddor, P.S. & Ohala, M. (eds.) (2007): *Experimental Approaches to Phonology*. Oxford: Oxford University Press.

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

[x] 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.

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

[x] 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.

[x] 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

[x] 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.

[x] 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

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

[x] 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.

[x] 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.

[x] 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.

[x] 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.

[x] Wagner, I. (1995): A new jitter algorithm to quantify hoarseness: an exploratory study. *Forensic Linguistics*, Vol. 1, No 1, 18-27.

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

[x] 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.

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

[x] 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.

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

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

[x] 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

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

[x] 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

[x] Whitmore, J. & Fisher, S. (1996): Speech during sustained operations. *Speech Communication* 20, 55-70.

16. AMENDMENTS AGAINST PREVIOUS VERSION

Not applicable (first version)

APPENDIX 1

Guidance for the Recording of Speech Samples for Forensic Speaker Comparison

1. General Remarks

The quality of the recording should allow auditory, linguistic and instrumental (F0, formant values etc.) analyses, i.e. most of the utterances of the speaker should be intelligible. Good acoustic quality (e.g. sufficient signal-to-noise ratio and a frequency range equivalent to at least telephone channel) is also required. The duration of the speech signal should exceed 5 minutes in total. It is also preferable to use a digital recording device. The recording should be made in an uncompressed raw format (e.g. PCM).

The speech of the person should not be monotonous, but as close as possible (in terms of emotion etc.) to the speech in the questioned recording. Therefore, it is strongly recommended that the individual who is responsible for recording the samples pre-listens to the questioned audio to ensure comparability (see §9.2). Therefore, it is necessary to achieve the greatest possible match between the circumstances of the known recording and those of the questioned audio. The technical characteristics of speech samples should be at least equal or better than the characteristics of the questioned audio.

Any voice disguise of the recorded person should be avoided.

All extraneous noises and interference (knocking, slamming the door, whispering, other voices, typing, phone calls, etc.) must be minimised. The recording should be made in a room with a low level of reverberation.

2. Procedure

The following steps are preferable:

1. First, the person should introduce themself (name, date and place of birth) and give a list of places of residence since birth (cities, regions, countries). Considerable periods, e.g. more than a year, should be mentioned. Then the person should tell an autobiographical story (in detail and as informal as possible).

2. Next, the person should read two pieces of text. Several readings of each piece are preferable (‘normal’, ‘as quick as possible’ and ‘emotional’). The first text should be a piece of fiction with a fairly simple vocabulary or, preferably, a phonetically representative text. The second text should be a (piece of) transcript of the conversation of the questioned recording.

3. Last, but very important for the following Forensic Speaker Comparison, the recording should include a natural dialogue with another individual. It might be a conversation about the case but could also be on any other topic. The purpose of this part is to get the samples in circumstances as close to natural speech production (dialogue).

All recorded speech samples should be saved on a data storage device (e.g. CD disk, memory card) and submitted (properly packaged and possibly encrypted) to the forensic speech and audio laboratory.