

Lecture notes- voice comparison

Voice Comparison Standards

Standardisation of forensic voice identification procedures and certification of examiners was commenced in 1971, when the Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification (IAI) were formed. IAI stopped their certification program in 1995, but since their standards of examination procedures are still valid, it will serve as the example in this paper. Other associations and organisations have partially or completely adopted this standard. The FBI's internal standard is very similar to IAI's and since this is the one used in one of the most important reliability references, comparisons between the two will be made regarding to where they are different.

Examiner qualification

The IAI's requirements for formal education of voice examiners are a High School Diploma with at least one additional course in speech science. A university degree with papers in related subjects such as electronic engineering, physics and linguistics are additionally recommended (required by the FBI). Every examiner must attend a short course in spectrographic interpretation before an apprenticeship period of usually two years commences. During this period the trainee is under the supervision of an examiner that has been certified for at least two years, with a minimum of one hundred comparisons in real forensic cases, resulting in at least twenty-five high confidence level conclusions. When this apprenticeship period is completed, the examiner goes through practical, theoretical and hearing tests every three years in order to be formally certified.

Reference Samples

The quality of the reference samples is critical in making an accurate analysis. Although these samples can be collected by the investigative officer, it is recommended that this is done by the examiner whenever possible. The quality of the unknown sample is dependent upon the recording conditions, which include such factors as the quality of the microphone, transition line and recording equipment. As these factors can to some extent alter the available frequency information, it is important that some of them are emulated in the reference samples. As an example: if the unknown voice sample is recorded over the telephone, this must also be replicated with the reference samples with approximately the same

distance between sender and receiver. The only things that must not be emulated in the reference samples are acoustic background noise and the quality of the recording equipment. For this purposes ²only quiet rooms and high quality tape recorders or digital sound systems are allowed. The speech samples, should ideally, be spoken in the same manner as the unknown voice. In order to achieve this it is recommended to have someone to recite the phrases in the same manner as the unknown sample, and have the suspect to repeat them in a similar fashion. If the suspect is cooperative it may be sufficient to have him/her to read in a normal voice from a familiar transcript. Unless the whole text is very long, every phrase in the entire text must be repeated until three satisfactory reference samples are obtained.

Aural Comparisons

Aural comparisons between the samples are performed as short-term memory examinations. This is done by having different playback units for each sample, with the ability to rapidly switch between them. The first things one will look for during aural comparison of the samples are huge inconsistencies such as sex and age that will exclude further examinations and result in a positive elimination of the suspect. In forensic examinations this is not likely to be a real issue, as the investigative officer on the case will perform this initial screening, and only the difficult cases will be submitted. When extreme dissimilarity does not exist, the examiner must carefully examine the samples and the individual words for components of speech that could be characteristic of the individual. The usual things to listen for during aural examinations are information about: pitch, intonation, stress/emphasis, rate of speech, mode of phonation, vocal quality and possible speech disorders. Pitch and intonation refer to the tone of the voice and the way in which this is altered through continuous speech. This can be useful in distinguishing between speakers, but the examiner must be aware that it is likely to be different if the samples are obtained through conversation or reading. These circumstances are also likely to influence other characteristic parameters, such as the rate of speech and the stress within a word or sentence. Obtaining information about these either from the officer who collected the samples, or purely through listening, are therefore essential. The presence of speech disorders and alternative modes of phonation will add important information to the identity of the speaker, as these can be highly characteristic of speech abnormalities. As mentioned earlier, the ear and the human brain have the excellent capability, not yet fully understood by science, to recognise different voices. The overall subjective judgement made by the examiner on the general similarity or dissimilarity between the samples will therefore also be of importance to the final conclusion.

Spectrographic Comparisons

3

A comparison between the reference sample spectrograms must first be performed in order to determine the range of intraspeaker variations for all the parameters. If there is a considerable amount of variation present in a specific word, this must be excluded from the analysis and the final conclusion. When only a limited amount of variation is present in the reference samples, the unknown sample must be examined to see if it is consistent with this range. The approximate frequency, bandwidth and relative intensity of the different formants, together with the shaping and positioning of the transitions must be examined for each word. One of the weaknesses with spectrograms is, as result of the trade off between frequency and time resolution, and is the fact that precise values of these parameters cannot be obtained. However, the purpose of the spectrogram is to analyse the shaping, the rate of change and the relative positioning of formants, something it displays perfectly. The energy distribution of the noised sounds must be compared to see if these are consistent between the samples. In samples of poor quality with only limited frequency ranges, this can be difficult as most noised sounds have important components above 4kHz. Similarities in fundamental frequency can be examined through the vertical striations present in the voiced sounds, as these will have timing correspondent to the vibrations of the vocal folds. This can be a very useful characteristic of a voice, but must be used with care when examining separate words, as it is normally altered continuously during intonation of continuous speech. The length of a word can also be characteristic of a voice and can be read off a spectrogram by measuring its length with a ruler. When a person speaks more slowly or faster than normal, it is usually the time between the words that is affected. The effects of inappropriate couplings between sounds can be seen in a spectrogram as either enhanced or diminished energy distribution in a specific frequency band. This is the result of irregular use of the resonating cavities, and results in highly characteristic properties such as nasalized speech. A good pattern match exists, when most, if not all of these parameters show strong similarities. It is important though to keep in mind that very different voices may exhibit similarities.

Expressing result

The IAI standards allow the examiner to produce one out of seven possible conclusions for each examination. These represent different confidence levels, which are determined by specific criteria. However, the examiner is allowed to choose a lower degree of confidence for a conclusion even if the criteria of a higher confidence level are met. An identification or elimination occurs when at least 90% of the comparable words are judged by the examiner as being very similar or dissimilar

through both aural and spectrographic comparisons. The total number of words must be at least twenty, and three or more formants must be visible. These decisions are not allowed if any of the samples contain vocal disguise or are recorded more than six years apart. A probable identification or elimination is allowed when at least 80% of the words are very similar or dissimilar both aurally and spectrographically. These must have a total number of at least fifteen words with two or more identifiable formants. If, in this case, only ten or more words can be produced the conclusion will be a possible identification or elimination. If the result of the comparisons falls below the criteria of the lowest confidence level the decision must be inconclusive. This must also be the conclusion if the samples reveal aural similarities and spectrographic dissimilarities, or vice versa. The FBI standard has the same criteria for the highest certainty level decisions, but do not allow any lower level decisions to be made unless there are at least twenty comparable words. In both standards the high confidence level decisions must be confirmed through the independent analysis of a second examiner.