www.easyvoicebiometrics.com
866 260 6376 Toll Free US

Info@easyvoicebiometrics.com
717 764 94240 Non US

Comparison Parameters and Speaker Similarity Coincidence Criteria:

The Easy Voice system uses two interrelating parameters of comparison (first and second error types).

False Rejection, FR is a probability of error when two voices are being compared that are considered different. The larger the FR value is, the higher is the probability that the two voices are of the same speaker.

False Acceptance, FA is a probability of error when two voice samples being compared are considered from the same speaker. The smaller the FA value is, the higher the probability that the two voice samples are from the same speaker

The result of comparing these two voice samples (one is a known and the other is an unknown voice) is the FR and FA values are calculated.

Likelihood Ratio LR is a measure of similarity between two voices being compared: it is calculated as FR and FA probability ratio. The larger the ratio value is, the higher the probability that the two voice samples are from the same speaker.

Comparing these three values, the system then calculates a generalized similarity value for biometric characteristics of the target and source speakers.

Values range from 0 to 100. The larger this value is, the higher the probability that the biometric characteristics are from the same voice and speaker.

The use of highly technical algorithms are employed which produce statistical numerical data that the program calculates and derives a false rejection number, a false acceptance number, and a likelihood ratio number and produces a unbiased result that it is or is not the same voice.

www.easyvoicebiometrics.com Info@easyvoicebiometrics.com
866 260 6376 Toll Free US 717 764 94240 Non US

The Easy Voice Biometric Analysis Algorithms:

The original analog Spectrograph generally referred to as the Voice Identification V700 or the Kay Sonograph was the tool of the day in which to perform the analysis. The technology at the time limited the analysis to the visual spectrograph comparisons of the formants, since the numerical information regarding pitch, rate, and other factors was not accurately available from the machines. Other equipment could be utilized to ascertain that information, but the listening and formant comparison was the main features on which a conclusion or opinion was formulated.

In the 1980's, Kay Elemetrics, with the help from the Russian Scientists at Speech Technology Center, produced a piece of software called Multispeech. This program was able to replace the analog machines and technology with a new digital version of speech analysis enabling the examiner to conduct the same aural-spectrographic method including all of the numerical results of the pitch, gaps, format tracking, etc.

So in essence, it was the start of the modern day Voice Biometric System. The examiner was able to identify or eliminate and make a verification of that analysis.

As a result of the Multispeech and IKAR Labs (Speech Technology Center) Voice Analysis programs, the "manual" method of Voice Identification requiring a verbatim exemplar has been superseded. Although it is advised whenever possible to obtain a verbatim exemplar, it is not always practical or possible. With the proper software techniques, a voice numerical model of a subject's speech can be obtained utilizing the original aural and visual cues listed above. This vast amount of information is fed as a model into a database and a detailed analysis can be conducted within the computer environment to discriminate between one model and another voice model.

The experience of the last 50 years in speech technology has taught us that the human voice is one of the best biometric descriptors that can be utilized

By using “model” criteria such as:

- Resampling of the signal to 11025 rate
- Determining the recording channel type
- Extraction of the speech signal (cleansing of the background noise and pauses)
- Calculation of the clear speech duration
- Calculation of the signal to noise ratio
- Calculation of the reverberation level
- Transformation of the signal files into a Riff Wav (.wav)

The ability to take these “models” and compare them in an “apples to apples” analysis goes a long way to ensuring scientifically reliable accurate results.

Voice Model Calculation:

In the current version of the Easy Voice (Voice Grid) three methods of biometric features extraction and pair wise comparison are implemented:

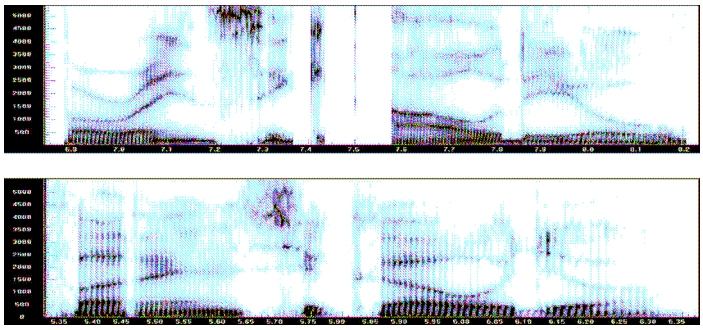
- Spectral Formant method.
- Pitch Statistics Analysis.
- Gaussian Mixture Models-based on Support Vector Machine (SVM) classifier.

www.easyvoicebiometrics.com Info@easyvoicebiometrics.com
866 260 6376 Toll Free US 717 764 94240 Non US

Spectral-Formant Method.

The spectral maxima of speech signal are called formants. They are formed because of the resonances which happen in the vocal tract during speech generation process. The formants (resonance frequencies) depend on the geometrical sizes and shape of vocal tract (head with all the cavities and organs). In general, in the frequency band of phone line (300-3400 Hz) we can find only four formants. The instantaneous values and dynamic traces of those four formants are extracted from the dynamic spectrogram and compared using Support Vector Machine (SVM) classifier.

In the picture below you can see the difference between the formant traces in the same phrase pronounced by different people.



Spectral-Formant Method provides high reliability of identification results and has the following advantages:

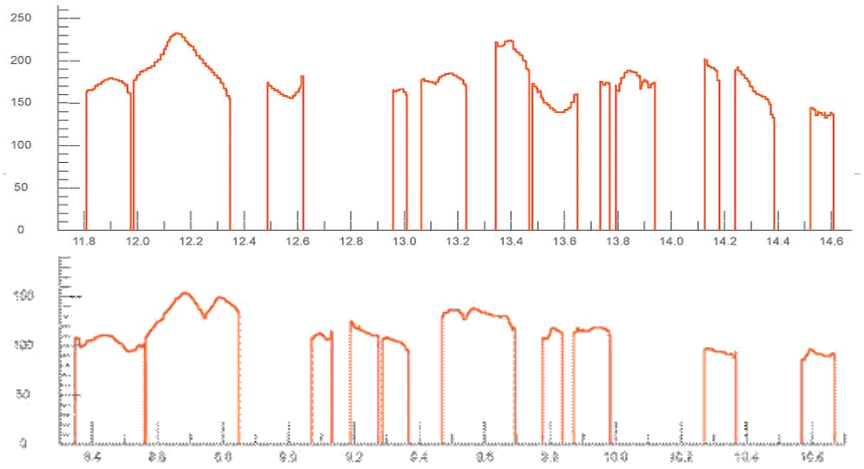
- Requires just as little as 16 seconds length of speech sample.
- Channel independent (channel features do not affect the speaker's model).
- Text and language independent and robust to emotional state changes.

• High noise-immunity features (signal-to-noise ratio as low as 10-12 dB). Noisy signals can be analyzed.

Pitch Statistics Analysis Method.

Pitch is a fundamental frequency of voice. It is the frequency of vocal folds oscillation. We can control and change this frequency (tone) depending on emotions and stress. That is why direct comparison of pitch value is not possible even for the same text. But the statistics of pitch are measured and compared. In the current version of Easy Voice) the 16 pitch parameters are analyzed:

On the picture below you can find an example of the same phrase pronounced by different people.

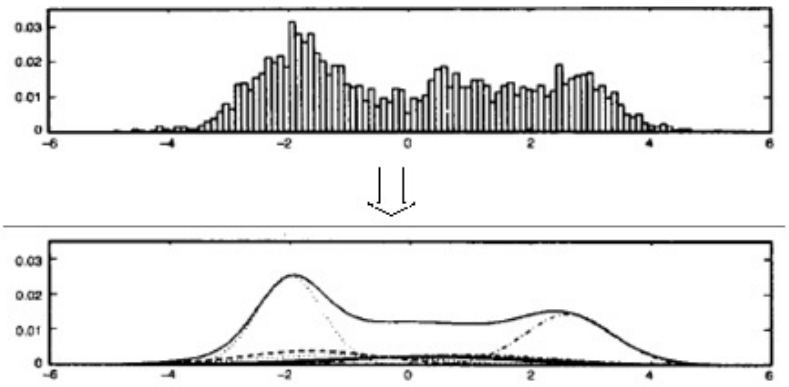


Pitch Statistics Analysis Method is an auxiliary method. The method provides lower reliability because of depending on an emotional state of the speaker (up to 16% of erroneous results are possible) and takes longer time to extract voice parameters and calculate speaker models. Still it shows the following advantages:

- Requires the minimum of 10 seconds length of speech sample (which is even less than the Spectral-Formant method).
- Noisy signals can be used.
- Channel independent.
- Small size of the model (about 1 KB per one sound file).
- High speed of the speaker search.

Gaussian Mixture Models based Method (GMM).

This approach is more statistical and requires computing power. So it cannot be accomplished manually. In simple words, not only the spectral maxima (values of resonance frequencies) are measured and compared, but the shape of those and energy distribution along the frequencies.



In another way, it describes the intra-speaker variability in one recording and compares it with intra-speaker variability on the second file.

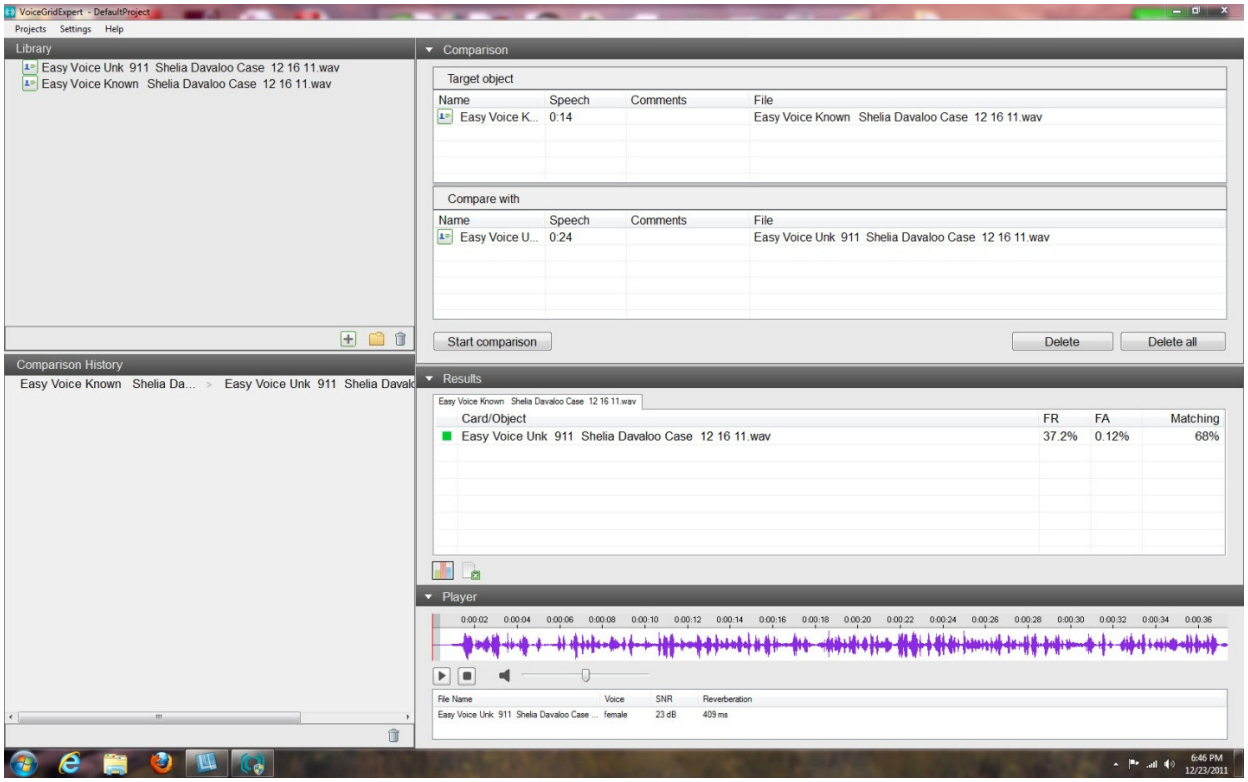
Gaussian Mixture Models based Method is the main comparison method. It demonstrates higher demands to the signal and duration. But it is the most precise approach. The highlights of the method are:

- High speed of the speaker search.
- Ideal for clear recordings with low noise level.
- Ideal for long recordings.

Results reflected in EVB

The green marker -111- indicates audio files which FR and FA values passed the strong Filter Thresholds.

The yellow marker-111- indicates audio files which FR and FA values passed the medium filter thresholds but have not reached the thresholds of the strong.

The grey marker -111- indicates audio files which FR and FA passed the soft filter thresholds but have not reached the thresholds of the medium filter thresholds.
The rest of the results will have no colored marker.

www.easyvoicebiometrics.com Info@easyvoicebiometrics.com
866 260 6376 Toll Free US 717 764 94240 Non US