International Journal of Criminal Investigation
Volume 2
Issue 1
57-67

ERRORS' LIMITS OF THE ACOUSTIC EXPERTISE

Saverio FORTUNATO
Specialist in Clinical Criminology, Professor, University of Lugano, Ludes.

Abstract

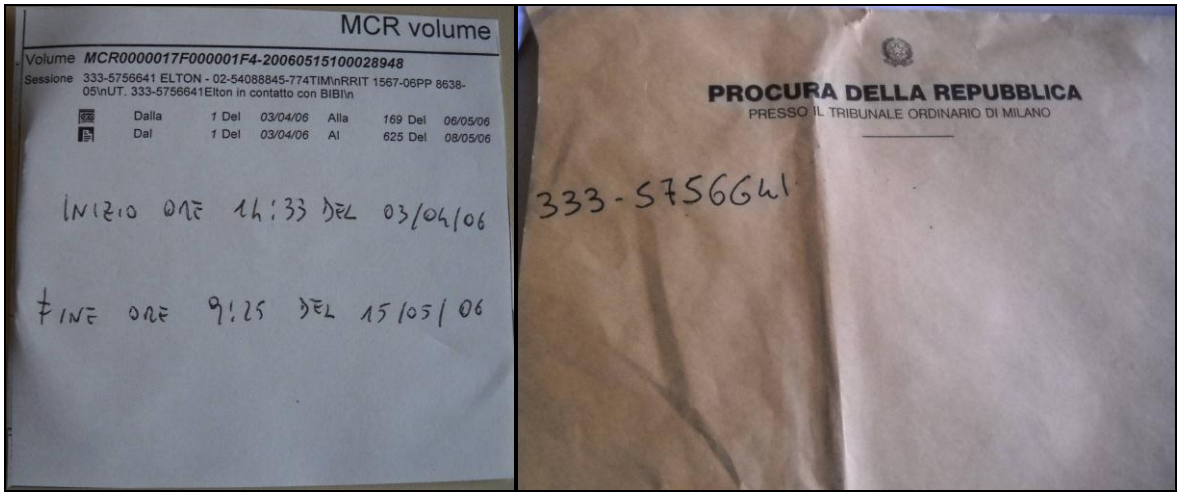
This study is concentrated on a very delicate subject, the one which concerns the acoustic expertise as a judicial prove. This type of expertise is based on the "voice print" examination. This term is subtly and deceiving, as it relates to other types of "traces" which have a very different meaning. When this voice print is spectrographically studied, it is observed that, intuitively, the highest tones will have a major activity of the high frequencies oscillations.

Keywords: *acoustic expertise, voice print, voice print examination, spectrographic analysis*

Introduction

In the context of a criminal trial held at the Court of Mantua (Italy), we received a mandate from a technical adviser, to answer the following question: « It is found that the scientific criteria of Angelo Giordano's report are the expected result both in method and in result ».

On 6.6.10 I went to Court in Ravenna, to access the case and to see the relative technical report on the interception made only of a number of users.



Technician position[1]

First of all, it is necessary to clarify that the "voice print" is a term subtly and deceiving, as it relates to other types of "traces" which have a very different meaning.

Digital fingerprints are, for example, mathematically speaking, simple functions and the theory is that if and only if there is an affine transformation bearing the imprint A overlapped on B, then the

[1] Dr. Leonardo Serni (Electrical Engineer), as illustrated by the author.

first speaker is the same with the second speaker.

Mathematically, an affine transformation is something very simple and if is limited to the axis of rotation and the purchasing problems are not taking into consideration (dust, fingerprints, dirt, damage, etc..), than the control is not very different conceptually of "window test": if overlap two traces and inspect them into the transparency it will see a unique trace, or two.

The calligraphic style is similar (parameters are different), but things get complicated because it is introduced the time factor.

Two digital traces taken at intervals of a year are substantially unchanged, except for changes in specific incidents. Two similar fingerprints, no, not quite.

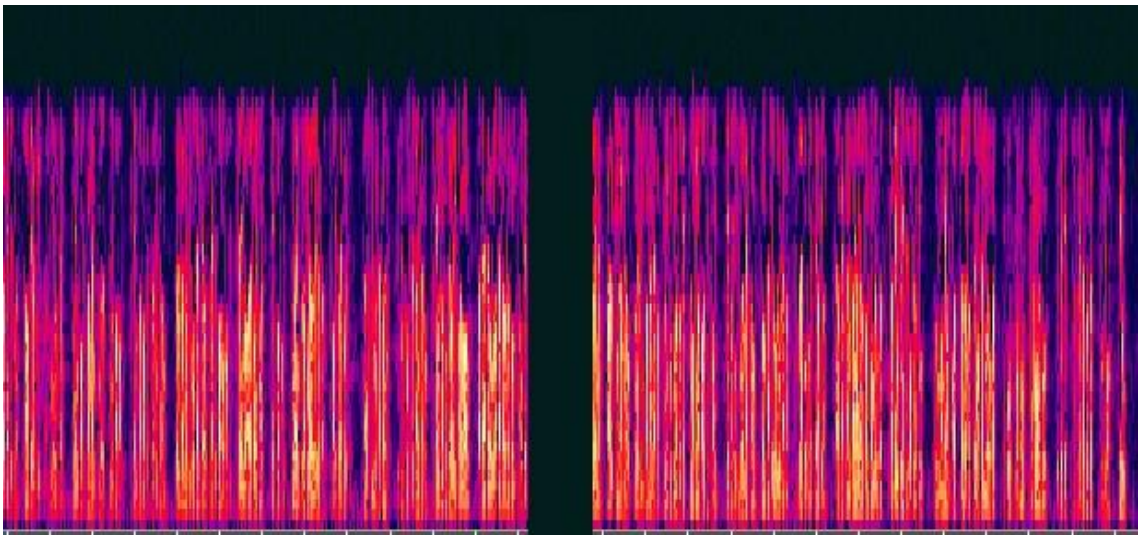
Printing voice presents the same difficulties, but more pronounced: the human voice system is controlled by several parameters, is more flexible than the neuromuscular system arm-hand (so much so that , for example, we are speaking faster than we can write).

It's like all these parameters are based or operating in soft tissues and muscles and the time dependence is much higher.

This does not mean that we do not recognize a voice the algorithms, but is much greater uncertainty.

It is an art quite complex, based on analysis of a series of parameters and especially on an estimate of "those" parameters that characterize a certain person beyond the temporal variations due to circumstances or improvisations (fatigue, emotion, arrochimento, setting voice. ...).

As an example we take a couple of vocal samples at random:



Here above, there is the voice spectrogram of Barack Obama and T. S. Eliot, a reporter from CNN while reading two poems.

The spectrographic representation is not very useful, so use a different mathematic tool, look at it by Fourier method (or as they say, "frequency domain").

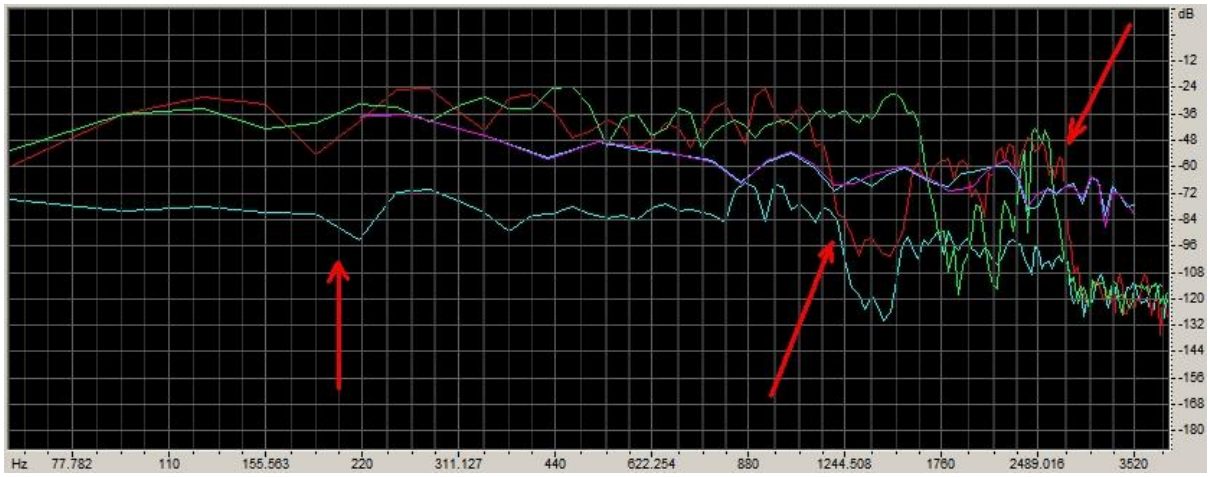
The idea is to assume that the voice is produced by a large number of oscillatory elements, such as diapason, some more active, others less and get to see exactly which and how many are most active.

Intuitively, the highest tones will have a major activity of the high frequencies oscillations.

Mechanically, the human voice is really produced by oscillatory phenomena, therefore this analysis works well; but they are not elementary, because we do not have a diapason in the throat and , in addition, the frequency of oscillation is variable (for example, easy stretching the muscles, their own frequency will increase and moves on fasetto.

In this review we see "spikes" corresponding oscillations upon those stronger and valleys in areas corresponding to the remote frequency corresponding the oscillations of string vowels (and harmonics produced by various resonances from the chest, throat, noise, and so on).

Some features "peaks" are indicated by arrows; the two Eliot's samples show up show the same ascents and falls (one is bigger and one smaller, because the two samples are taken at slightly different volumes), Obama's show others and then you might say "the red and cyan samples are samples from the same person" ... but look better and you will see that red and cyan are not IDENTICAL effects everywhere. Indeed.



You can see now the fourth diagram, represented in green. Is still T.S. Eliot while speaks normally. "Slope" which was before about 1 250 Hz is now

of 1760 Hz. The fall of 2500 Hz is always the same and is flanked by red and green all over in the range 70-1200 Hz, but with peaks of synchronization.

Since we know that T.S. Eliot is the one who always speaks, we can conclude that a change of the peak matters little, otherwise we would say that the green and red are two different people (but he's always Eliot).

The parallelism should count less, "shape"-in fact, we can overlap the green, red and cyan using a simple linear transformation, while with Magenta (Obama) does not work.

But even the "form" is not always the same, even for the same person: because if I had a sample of Eliot's while she speaks with the bartender at the bar would be different if I had a solemn voice reading the Water of Death. At the phone will also be different (we should seek other sources of similarity) and so on.

The **contrary** problem sounds like this: "These two are one and the same person?" and is even more difficult because of all the parameters that you can use, many of these do not depend on the person concerned-by its uniqueness, its horismos – but only accidentally : in body size, social class, wealth, status, even the level of alcohol in the blood.

Then we shall see two different spectrograms with five or six peaks perfectly identical and say: it is the same person; but in fact no, we could have identified the characteristic peaks of the speech in Missouri, and from a group of people who speak and that would fit these tips there cannot be found even one: all the men of the same age and the same size of the body came from the province of Saint Louis.

I repeat, there's not a thing that you cannot do; the problem is that in order to overcome these problems, it takes a lot of data, by a very large sample; and then the data in the best position to be comparable (not when Eliot who speaks on the phone compared to Eliot who reads poetry, but, if possible, even the same poetry reading Eliot). My opinion is that this is still much, much more an art than a science.

Introduction to the analysis of noise

1 - Voice recognition in the sense of the Commission the Project RISE

The Commission the Project RISE has decided over the voice recognition system. Dr. Francesca Mancuso, in her report of 27.4.2009, named "Project RISE, biometrics between ethics and safety", illustrated: «Biometric Systems based on the voice recognition takes into account

some features of spoken language, such as speed, frequency, structure and density of sound waves. The technique is based on using a microphone and software for the development and processing data.

The possibility of errors is great because the voice and manner of speaking are related to many physiological and heritability factors such as introducing continuous changing. Just as an example, a cold or an offense may impair the voice features. For this reason the voice recognition must be accompanied by other biometric techniques. »[2].

2 - Speech recognition in the University of Cambridge

Prof. W. J. Barry, CE Hoequist and F.J.Nolan from the Department of Linguistics, University of Cambridge, United Kingdom, says: In Automatic Speech Recognition, the methodical consists in "to take account of the systematic difference of voice depending on geographical location, regional, local". In their laboratory, they split the task into two phases: during the first stage, they adopted a procedure for identifying of accent, selecting one of the four English regional accents according to differences of speech quality within four calibrated sentences. In the second step, the procedure for registration moved the voice field from the field of the regional criteria to the speaker, the vocal field being exactly as it is calculated by the information of the identification of the accent. All of these are used to highlight the problem of interaction, the regional accent, pronunciation or speech defect and so on and all the diffcalt variables, if not impossible to classify more or less with an absolute certitude.

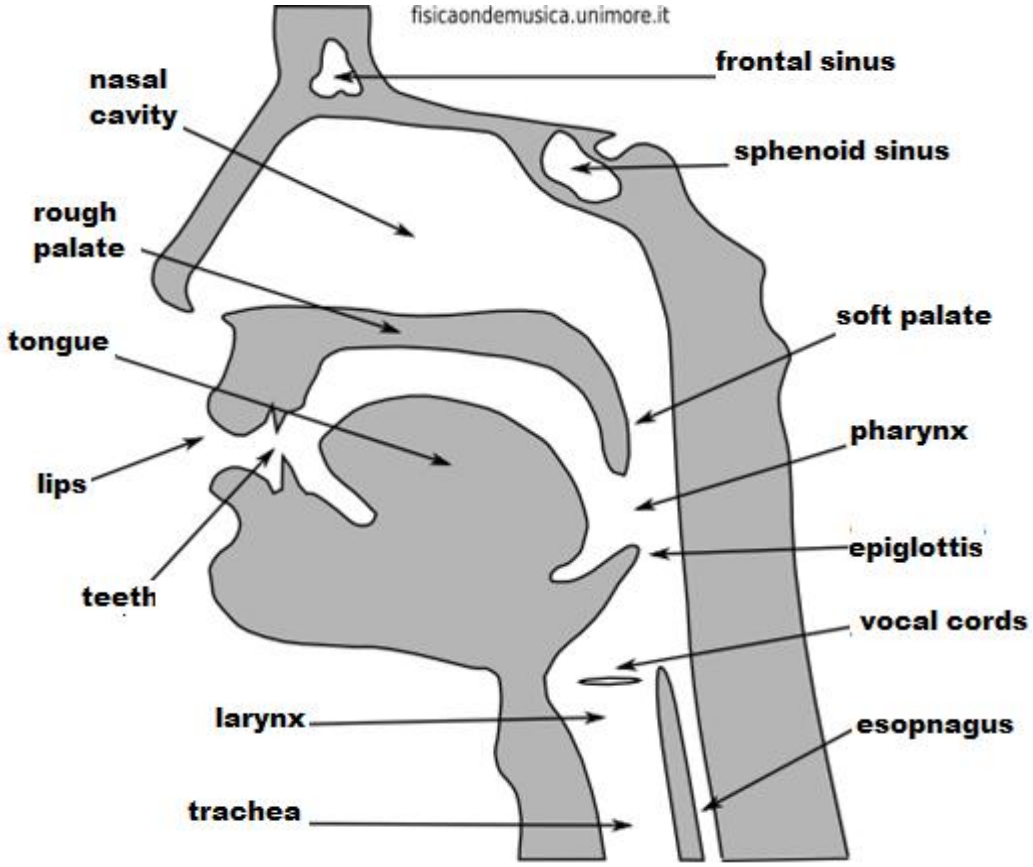
The human voice is not a vocal phenomenon

CTU defines "auditory phenomenon" the human voice and it is assimilation with the generation of sounds in a musical instrument. But this comparison is approximate, because the human voice to generate itself requires: air (which comes out at a single press of the diaphragm), from the vocal cords whose thickness, length and voltage determines the height of the sound, thoracic cavity with the mouth, nasal cavity and various other small areas, distributed in the cranium.

The particularity of the human voice exist in that the mouth is mobile and this allows to vary the sounds form emission. It follows the possibility of variation of vocal timbre with continuity

[2] Source: Scienzenews.it

(e.g., maintaining of a constant height, moving from "a" to "u") but that is not automatically included in the musical instruments, because the particularity of being able to vary the timbre of the voice is similar to its transformation, without continuity, from the sound of a clarinet in that of a flute.



Simplified diagram of the human vocal tract, by Carlo Andrea Rozzi

The identity of the voice cannot be probabilistic

At p. 5 Mr. Giordano writes:

Voice expertise cannot provide an answer in terms of probability. A claim of type: ***«calculations performed on the voice signal provide exact at that time, but only indirect and deductive data of the speaker, for a final analysis of the voice's tasks being considered only of probabilities type»*** is vague. If the answer to your question would have an outcome highly probable, then it would be sufficient the manual statistics, putting two tickets in a bag, writing on them true or false and pulling at random: we will have a 50% chance that result to be exact.

It is very obvious, that it is not necessary the expert to obtain a probability but a reliable opinion that is away from

error. At the Lottery, in a rough way, it is believed that recent newcomers have a different probability of departure. In reality, however, that plays a number of income among the last, it has the same probability of escape as any another number. Then, the answer given to the analysis even in terms of "another probability", on scientific grounds, it's a non-response for that leave the problem unresolved and stop attempts of the expert from the beginning.

The identity of the voice could not be based on the auditory perception

In terms of perception, the philosopher **Melisso**, disciple of Parmenides, notes that in reality, what we perceive through our senses (sight, hearing, smell, touch, taste) would be valid only if the things collected would always be like the first time we've seen them. Then, even most of the things need to stay the same. This means, that the existence of multiplicity will be accepted only if there will be similar characteristics of unity. Instead, Melisso say, owe all say, all of us must be taken account of the fact that the things which we perceive being sensible things in the world of sensitive, things from the empirical world are never the same, but they continue to change, because they are born, live, die, break, they do disappear, etc.

Therefore, there is a profound contradiction between what is true and what, however, our senses show[3].

For **Arthur Schopenhauer**[4] the problem of cognoscibility was this: «What we can perceive using the eye, the ear, the hand is not intuition: There are simple data. Only after the intellect goes through back from effect to cause, the world may exist, as well as intuition into space, changing its form secondary, permanent and eternal in the matter itself […] the world whose performance is only in intellectual environment exists only through the intellect»[5]

All these things exist in order to have to be able to say that the phonic analysis is inevitable used for something questionable, suggestive, where perception itself is something that changes the state of consciousness, depending on how prepared

[3] *« If, however, the existence exists, it must be unique. And being unique, must not have a body [...] In fact, if there is a body, would have parts and then there would not be only one». [Simplicio, Physica, 29, 22; 109, 20]*.

[4] Arthur Schopenhauer was born in 1788 in Danzica, into a family of wealthy traders. On the death of his father (supposedly by suicide) leaves with his mother for Weimar. Legacy left him by his father allows him to study. In 1809 enrolls at the Faculty of Medicine, but three years later he switched to the Faculty of Philosophy, graduating in 1813 with the thesis called *Sulla quadruplice del principio di radice ragion sufficiente*.

[5] A. Schopenhauer, *Il mondo come volontà e rappresentazione,* I, 4, translated by N. Palanga, edited by G. Riconda, Mursia, Milano 1982.

we are to perceive phenomena that we observe with the help of the senses (in this case hearing which means listening to a voice).

On the basis of the theory Mamona and Assaleh Logefoged

These researchers say that instead of using different parameters of the voices' spectral components, it is chosen a person. In this case, Mammone and Assaleh, use "*weights*" to determine the percentage of the speaker identification.

Logefoged says: "A decisive step in computed identification of the individual characteristics contained and in the voice formant would be offered the possibility of inserting the voice sounds in a machine or a computer to obtain a transcript with phonetic accuracy (with all shades described by the symbols contained in the phonetic alphabet)".

In reality, the phonetic alphabet is already a simplification, for example notes from a score.

Two different speakers could hand down the same sound in different ways; is the case of appropriate symbols |dζ| |θ|. Or it is believed that the vowel "a" as in the city of Bari in Italy becomes "e" (as is well known in Italy, people of Bari pronounce Beri, instead of Bari).

The voice identification cannot be realized in compatibility conditions

Argue that the comparison of voices elements which are assumed to belong to the same investigated suspect shows "*the degree of similarity*" or compatibility and use the expressions such as:" *There is compatibility encountered relative to the fundamental frequency of speech in similar contexts at the same rate of speech*", it has nothing to do with science.

The reasoning is false, for that is equivalent to saying that "part A is complementary to part B". In reality A and B are part of a greater whole. (B) include, for example, all Albanians, of the age x not y, with a mass c not d, etc. It's like it is used the sentence

"*My oil is good because when where it is fried there is no smoke*", it says something true but in particular no oil do not smoke when it is fried, so my oil is not a good one but is an oil as all the others.

The result is true but the impact is false. It is true the result but only if it is used by exclusion (not included; it is used the argument that the "oil is not good". Even Hollien says (referring to the voice recognition) that the method is good through exclusion not through inclusion: when you say that the voice is belonging

to Tizio, it is well; when it is used to say that is Tizio, than the method is false.

Serious error in comparison process

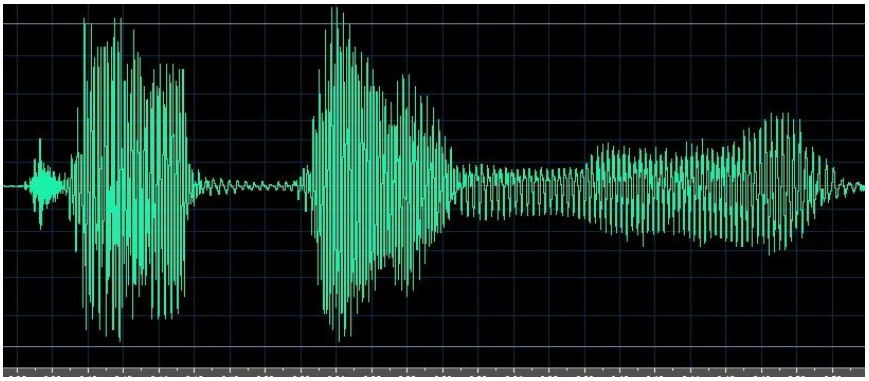
The comparison is usually made from a voice of the suspect during interrogation with the existing recorded voices called (rightly) "unknown" and suspected to belong to the suspect. Voice during interrogation is common and spontaneous influenced by the moments of solemn silence of the interrogation itself; unknown voice will be conditioned by that particular moment when the subject spoke.

If it is true as it is, the voice is composed of variables such as excitement, fear, etc. than how can it can be made the comparing of the specific voices and conditioned by the context of the past in order to get a similar result to that date for an ID?

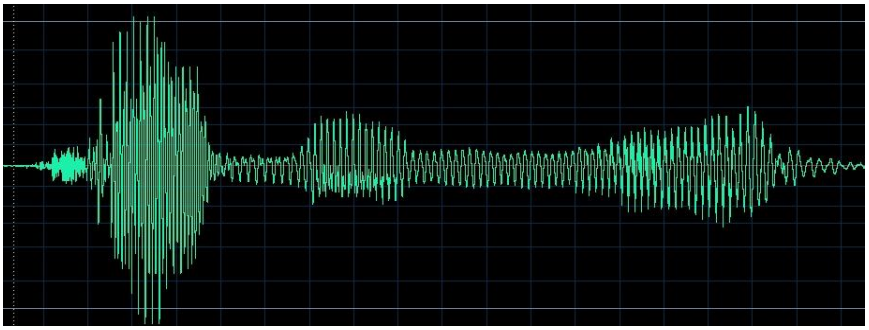
Proof of noise analysis error

As proof of his epistemological noise evaluation expertise, I describe the following experiment: with the help of ing. Serni, I chose incidentally, by listening, the recording of the process with the No. 12, of identifying a word (the original word, from the recording is in Albanian language) uttered twice within the same telephone conversation.

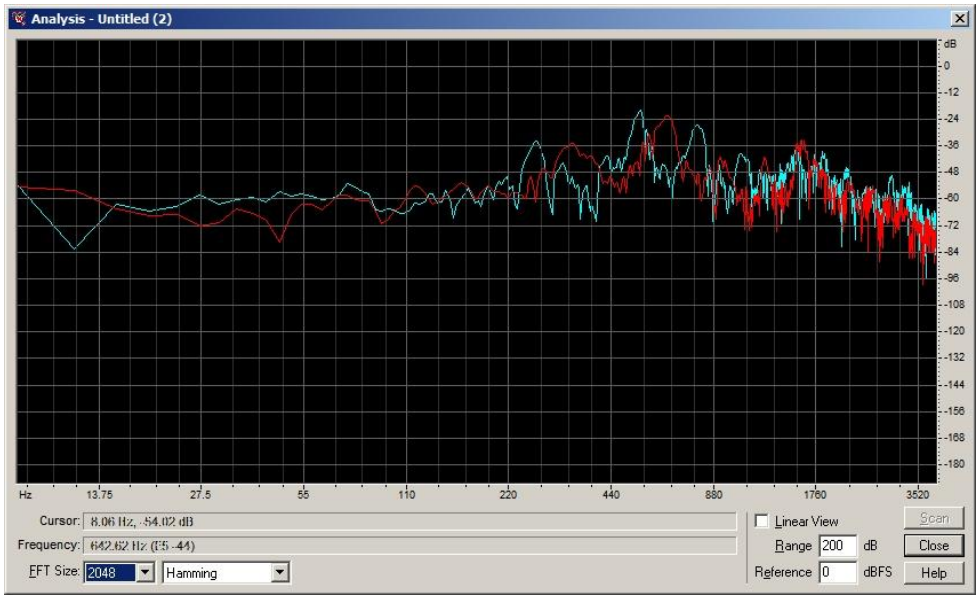
By comparison, the analysis revealed that this word belongs rather to another person.



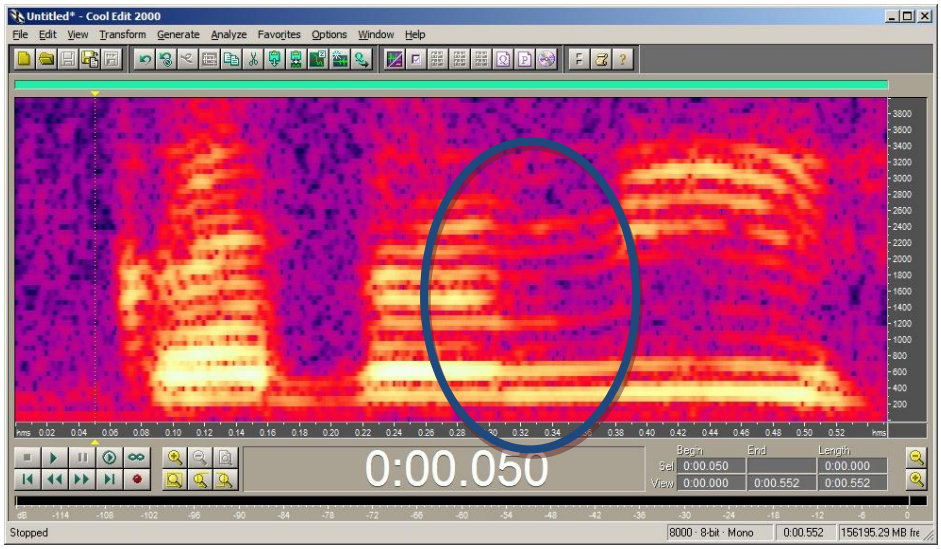
Above, 1 ° the first fragment of a word said for the first time in the telephone call 12



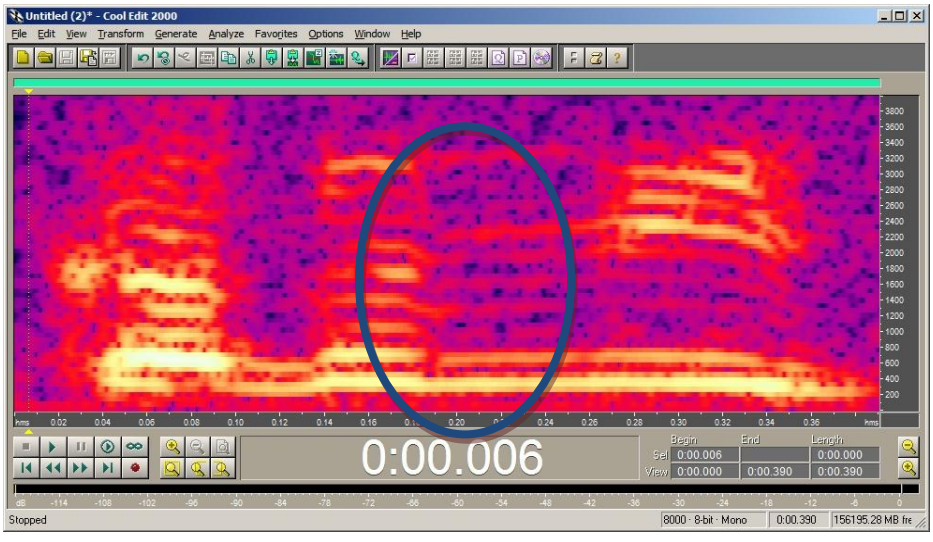
Above, the second fragment of the same word uttered in telephone call 12



As noted, there are the same points, but it is a incorrect result of speaker's diversity.



Spectrogram of the analyzed voice



It may be observed the spectrogram diversity in the area cordoned off, showing a variety of vocal styles and translated into reality is about the same speaker in the same call where the speaking a word twice is the same at a distance of a few seconds to another.

The result: voice analysis works well only if it is used by excluding the vowels identity; is not a scientific method including, meaning the identification or identity of the voice. It can be said that voice does not belong to Tom, but it cannot be said that it is Tom.