DOI: 10.17469/O2110AISV000007

SONIA CENCESCHI, CHIARA MELUZZI

Transcription and voice comparison of noisy interceptions: remarks from an audio forensics report[1]

The paper describes a case study of particular interest and representative of speaker identification and speech attribution problems in real environments. The authors were recruited as technical consultants by the accused (already convicted) person's lawyer to report on a probable misidentification of her client in two environmental recordings, with the goal of obtaining a scientific analysis and subsequently requesting an eventual review of the sentence. The results, carried out by the experts blind to the overall criminal proceeding, led to a conclusion of rare clarity in this field, and to the evident presence of a clear mistake in the speaker's voice attribution. It is believed that the description of the work (albeit anonymized) may be of strong interest to those working for the promulgation of audio forensics and forensic phonetics scientific methodologies in real contexts such as preliminary investigations and legal proceedings.

Keywords: Forensic phonetics, Audio forensic, Noisy recordings, Speech perception, Psychoacoustics, Speaker comparison, Speaker recognition.

1. *Introduction*

Noisy audio has usually been disregarded by phonetic analysis, especially those dealing with the quantitative comparison of acoustic features extracted from different speech samples. To eliminate interferences, one of the first things researchers learn is to record in a silent environment with no technological devices turned on. Furthermore, audio files have to be recorded in WAVE (.wav) format with a sampling frequency of possibly 44.1 kHz and 16 bit (Di Paolo & Yaeger-Dror 2011, Meluzzi 2022). When working with forensic audio files, the situation is rather different, both in terms of audio quality (i.e., recording format, noise, and so on) and the amount of acoustic data available for a proper and reliable analysis. What happens in forensic cases is that a person is under trial on the basis of many proofs, thus including one or more audio files. These audios are usually intercepted recordings made through different technologies such as micro audio recorders or trojan, and more recently WhatsApp audio and similar kind of social network audio messages (for a full case report see Cenceschi, Meluzzi & Nese, 2020).

[1] This work has been jointly conceived and written by the two authors. However, for the Italian evaluation system, Chiara Meluzzi is responsible for section 1, 3, 3.1, 4 and 5; Sonia Cenceschi is responsible for sections 2, 2.1 and 3.2.

This difference between the 'optimal' audio data expected in phonetic analysis, and the forensic reality is the main cause of misunderstanding in expectations between the phonetic expert and the so-called 'third part', e.g., the judge, attorney or law enforcement officers asking for a forensic report. On the one hand, this implies that the experts have to be very clear in their report: not only regarding which methods they have used, but also which have to be excluded and on what scientific basis. For instance, a statistical analysis can be run only if some pre-requisites of the matrix are fulfilled. To begin with, there must be sufficient data in all the conditions to be tested: if the intercepted file is too short, sometimes less than 10", and contains very few speech samples from the subject to be tested, it is impossible to provide more than an opinion, albeit well-grounded and justified by descriptive evaluations based on specific knowledge. On the other hand, third parties usually expect a clear Yes-or-No answer, for example, from a phonetic comparison, but this could rarely be asserted without a margin of doubt. Thus, a forensic expert needs to balance the scientific reliability with the practical needs by using a lexicon and examples understandable by naïve users (i.e., not experts). However, without tangible experience in real-world forensic circumstances, these dynamics and balance are impossible to achieve or discover.

In this paper, we will present a practical working scenario of a real-world example on which the authors have worked. For privacy reasons, the name and surname of the person involved in the trial, as well as other information that could identify him have been avoided, by referring to him as 'Mr. G'. In the course of his trial, Mr. G was accused and convicted of a crime based on two intercepted audio files. Mr. G's counsel wanted to know whether Mr. G. was correct in claiming his extraneousness to the facts in order to proceed with a judicial review motion. We, hereby, describe the case study in detail (section 2), including the materials, the methodological choices and the important remarks inferred and collected during the work progress. Section 3 contains analyses for both of the attorney's requests, and the report's findings are summarized and discussed in section 4 to highlight the issues in working with noisy audio recordings in forensic phonetics, as well as the role of forensic experts in balancing scientific reliability and practical needs. A final section focuses on the conclusion and future possibilities.

2. *Case study: audio comparison in Mr. G.'s trial*

After the first trial was over, audio forensic experts were called in to help with the case. Since Mr. G. was asking for a judicial review, he asked his lawyer for an audio forensic report on two intercepted environmental recordings, which consisted the unique incriminating evidence. There was no additional technical information available, as it is typical in such circumstances (e.g., placing of the microphones, presence of external sources of noise). The transcription of a recording done by Law Enforcement Officers during the preliminary phases of the investigation, reported the surname of Mr. G. as pronounced by an unknown voice. A second recording

comprised a short excerpt of the speech that the police officers matched with Mr. G.'s speech timbre. On these basis, Mr. G. was prosecuted and then arrested, being later sentenced to 18 months in prison. He spent these months at house arrests, by always claiming his innocence. Only after 10 months, the Mr. G's lawyer obtained these audio recording and asked for an expertise report.

In June 2018, the two forensic specialists were called and asked to provide an audio forensic report for two main purposes, that is to determine:

1. whether the first environmental recording contained the surname of Mr. G.;
2. whether the second environmental recording contained the Mr. G.' speech.

The first point is concerned with perceptual, spectral, and phonetic observations, whereas the second is a classic comparison between a Known Voice (KV), that is, the client's voice as recorded by the forensic experts, and an Unknown Voice (UV), as reported by Law Enforcement Officers and identified as the client's voice.

2.1 Materials and data

Mr. G's lawyer provided the audio recordings to the forensic experts for their report. Table 1 summarizes the data, by dividing the audio belonging to the Unknown Voice (UV) and those pertaining to the voice of the client (Known Voice, KV). The table also reports an indication of the duration of the original audio file, its original saving format, and the typology of recording (i.e., environmental interception, phone call, etc.). No information was available concerning the instruments used in the interception.

Table 1 - *The data available for the report, as provided by Mr. G's lawyer*

Typology	*Total duration*	*Format*
Environmental interception With the Unknown Voice (UV)	19'06"	.ogg
Environmental interception containing the surname	19'40"	.ogg
Phone Call, spontaneous speech (Known Voice)	5'37"	.m4a
WhatsApp, spontaneous speech (Known Voice)	1'49"	.m4a
WhatsApp, spontaneous speech (Known Voice)	42"	.m4a
WhatsApp, read speech (Known Voice)	1'	.m4a

The presence of more data for the KV is normal during a phonetic forensic comparison, especially if the KV belongs to the lawyer's client and/or a person under trial. In these cases, the forensic expert could pretend to record more audio samples for the comparison by performing the so-called *saggio fonico*, that is a collected audio sample from speaker comparison (cf. Kersta, 1962). Obviously, what determines the possibility of effectively performing an audio comparison is not the amount of data

at disposal for the KV, but what we have for the UV. In the case presented here, the only audio files belonging to UV to be compared with Mr. G. was an environmental interception recorded in a public domain compressed audio format (.ogg). Despite its length, the audio was very noisy and different voices overlapped with each other. This means that the amount of speech effectively produced by UV without overlapping was very poor both from a qualitative and quantitative point of view.

The *ogg files were in stereo format, 16,000 Hz, 16 Bit, and had a useful upper spectral threshold of 8,000 Hz. They were converted to wav format (dummy enhancement) for software processing convenience. Although they were of inferior quality, they had a threshold SNR (signal to noise ratio) sufficient to perform spectral and phonetic analysis with a semi-automatic method (SNR > 60 dB). The materials belonging to KV consist in:

1. a business call from Mr. G;
2. three vocal messages recorded on WhatsApp[2] by Mr. G and sent to the experts. The audio files were recorded with a Huawei P30 PRO, with a sampling frequency of 48 kHz, 16 bit. Two of them contained spontaneous speech, and one WhatsApp audio message for which Mr. G. was asked to read the same linguistic content pronounced by the UV in the environmental recording.

3. *Analysis of Mr. G's case*

A forensic report must be clearly organized and convey the results in the most exact but simple manner possible in order to improve the understanding of specialized aspects by so-called "third parts" (i.e., judges, lawyers, etc.). The necessity to maintain a dialogue between the expert and not-expert comes with the requirement to explain in detail the scientific jargon and the different tools normally used in phonetic analysis. It was important, for example, to explain what a spectrogram was, how PRAAT was used, and to provide a brief summary of the results at the beginning. For clarity, results were here organized in the same manner: A first section (3.1) is dedicated to determining the presence or absence of the client's surname in the audio files, and a second section (3.2) is dedicated to demonstrating whether KV and UV could conceivably belong to the same speaker (i.e., Mr. G.).

[2] WhatsApp is based on the SILK VoIP (Voice over Internet Protocol) protocol, a codec developed by Skype and now license-free, available in open-source mode. This technology allows you to have a conversation similar to that of a telephone network by using an Internet connection or any other dedicated packet switching telecommunications network that uses the IP protocol for data transport. SIL VoIP is the basis, with CELT, of the hybrid Opus codec, the official format of WhatsApp, which is a lossy audio coding format used to achieve low latency with the best quality. In particular, WhatsApp exports Opus files as m4a (with the AAC codec) because Opus is not supported by many audio player applications.

3.1 The presence/absence of the surname

In order to preserve the anonymity of the client, in this section we limit to describe the procedures adopted in the report, without providing the examples included in the report to demonstrate the presence of a different surname than the client's one.

Following pre-processing, the first stage was the extraction of the portion of interest from the original *.ogg file, which was already been documented as part of the ongoing legal procedure. The portion of interest corresponded to 13 seconds (from 5' to 5'13" of the original file). It was saved as a separate file called "Extracted V1". The file was then converted in .wav format to allow formants' viewing in PRAAT[3]. The new file was transcribed and time-aligned with the help of ELAN software (Lausberg & Sloetjes, 2009) in order to annotate different linguistic and extra-linguistic features on multiple tiers. Only the transcription tier was then exported in a Word file to be attached to the final report. We reported below in Example 1 an anonymized transcription of this file: proper names and surnames have been substituted as XXX and YYY, respectively. The amount of Xs and Ys correspond to the original phones heard during transcription. Voice 1 and Voice 2 indicates the two (male) speakers. The English translation for each line of dialogue is provided below; this was not part of the original report because it was unnecessary.

Example 1: Anonymized transcription of Extracted V1

1 Voce 1: c'è XXXX, c'è... ...YY(Y) [non chiaro, voci sovrapposte]
Voice 1: there's XXXX, there's... ...YY(Y) [unclear, overlapping voices]
2 Voce 2: XXXX?
Voice 2: XXXX?
3 Voce 2: chiama(lo) [non chiaro, voci sovrapposte]
Voice 2: call him [unclear, overlapping voices]
4 Voce 1: c'è XXXXXXXXX
Voice 1: there's XXXXXXXXX
5 Voce 2: YYYYY?
Voice 2: YYYYY?
6 Voce 1: sì.
Voice 2: Yes.

According to the excerpt, we were dealing with a very short dialogue between two people, the speakers of which were lately identified based on another audio file provided by the lawyer. The audio quality was very low, thus in *line 1 and 3*, it was specified that the content was unclear and that the two voices were overlapping too much to allow a correct transcription. Although this is not a proper scientific way to indicate overlapping or noise in a transcription (see, for instance, the guideline for transcription in Conversation Analysis; cf., for Italian, Savy 2005), it was necessary

[3] PRAAT now allows also .mp3 files, but at the time of the case it was not possible, and it was thus decided to convert the .mp3 to .wav only for visualization purposes. The authors discourage from extracting acoustic values from originally compressed audio files.

to provide a transcription as clear as possible to the lawyer. Thus, the use of symbols had to be limited or even avoided to enhance the communication with the third part.

For the purpose of the report, we focused on *line 5*, in which the second speaker pronounced the surname of the person that the prosecutor has identified with the client. The transcription already pointed out that this surname was constituted by 2 syllables and 5 phones, with only a small doubt on the third one, which could alternatively be /l/ or /r/. An examination of the spectrogram using PRAAT validated this transcription (Boersma, 2001).

In the report, pictures of the wave form and spectrogram as visualized in PRAAT were included. Each phone was delimited by vertical yellow bars and clearly indicated by an arrow. The explanation of the acoustic nature of each sound was also provided in the text, with a long note containing the major references to the topic.

In the end, it was possible to argue without doubt that the surname pronounced in line 5 of the excerpt (1) clearly show the presence of only two vowels on the spectrogram, thus corresponding to a two-syllables word. The phones constituting the two syllables could also be easily identified, with the only exception of a sound at the end of the first syllable that looked like an [l], but that could also be a rhotic realized as an approximant (cf. Celata et al. 2016 for spectrographic characteristics of Italian rhotic realizations). No fricatives or palatal sounds were detected on the spectrogram.

Since Mr. G.'s surname was four-syllables long, and it contained a fricative and a palatal nasal, it was possible to sustain the claim that in line 5 of the excerpt, the second speaker did not pronounce the client's surname. For this reason, the report claimed that the original transcript was incorrect. The hypothesis was also floated that Mr. G.'s surname was mistakenly added to the original transcript due to a perceptual error. Indeed, police officers frequently do transcriptions and recordings without the correct audio equipment, which can easily lead to blunders such as the one in this case. In this regard, it should be highlighted that the two specialists first examined the spectrograms and then listened to the audio to avoid being swayed by their knowledge of Mr. G's surname. Even so, the listening of the audio with closed headphones clearly confirmed the result achieved through the spectrographic analysis.

3.2 Speakers' voices comparison

Since the two audio recordings to be compared were both severely damaged in terms of quality and were also relatively brief, it was determined that no semi-automatic speech comparison could be performed. Nevertheless, we aimed at providing a reliable report by comparing the two voice across three different methods:

1. Perceptive test
2. Phonetic and voice quality analysis
3. Sociophonetic profile of the two voices

The perceptive test was conducted on three distinct categories of listeners: a group of 6 trained Italian phoneticians, experts in audio and/or forensic analysis; 6 foreigner phonetic experts, with different L1s and only a superficial knowledge of Italian; a group of 43 naïve listeners (24 women, 19 men), aged between 23 and 64 years, and

native speakers of North-Western Italy. The reason for testing this three different groups was twofold: on the one hand, we needed the opinion of trained experts who had previously worked with noised audio files for forensic purposes, but we wanted to see if the noised audio file's content could alter the experts' impression (e.g., Fraser 2003). On the other hand, we sought to compare the experts' results to the perception of naive speakers, to see if knowledge of the task and its repercussions affected the judgment on the similarities between the two voices.

Each group had to listen to two stimuli previously prepared by the researchers for maximum two times. After the listening, participants had to indicate if they believe the two voices in the stimulus belonged to the same person, the degree of confidence of their judgment and to eventually indicate what has shaped their judgment. Both stimuli contained two audio excerpts of around 10" each, with an interval of 3" of silence between them. Stimulus A contained the UV as extracted from the materials provided by the lawyer and a sample of KV as extracted from a phone call made by the client to his lawyer. Stimulus B contained the same two files but in reversed order (KV followed by UV). Each respondents started the test alternatively with stimulus A or B, in a random order. All groups of listeners agreed that KV and UV belonged to two separate people. Only 9 naive listeners out of 43 total respondents claimed that KV and UV belonged to the same individual, even though they mentioned in the open remarks that the accent and 'speech mode' were different. Comments from naïve listeners' provided useful information on the phonetic and phonological characteristics that have shaped their perception: possible geographic origin, as carried by regional accent, and speech rate were the most common features, together with the pronunciation of the vowel /e/.

A detailed phonetic analysis was provided by extracting the first formants' values of all vowels produced in the recordings containing both KV and UV. Table 2 shows the amount of vowels considered for each voice. Although usually only stressed vowels are considered in phonetic analysis, in this case (as well as in other forensic cases) the scarcity of the audio material forced the experts to use all available vowels, regardless of the stress position. Only diphthongs and vowels included in ruined audio segments were excluded from the analysis. Vowel /ɛ/ was also excluded because of the very rare occurrences in the data.

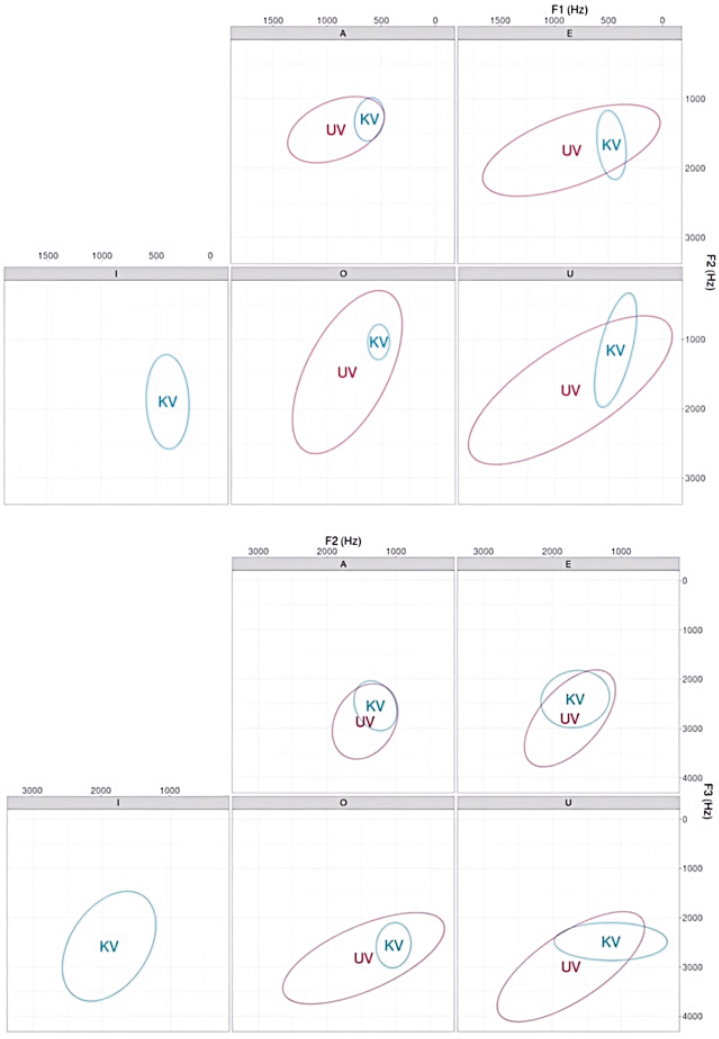
Table 2 - *Vowels available for phonetic comparison of KV and UV*

	Unknown Voice (UV)	*Known Voice (KV)*
/a/	67	60
/e/	84	64
/o/	17	43
/u/	33	15
/i/	0	32

The data were unbalanced, as it usually happens in forensic research, and many items had to be discarded because of the low quality of the audio file. We focused the analysis

on F1, F2 and F3 values, since F0 was too variable, and it could also be affected by the different registers and speech styles of the two recordings. Then, we plotted the values through Visible Vowels (Heeringa and Van de Velde, 2017) to enhance the clarity of the results for the third part, since graphical representation helps the understanding of reports. We plotted the vowel space for F1xF2 (Fig. 1 above) and for F2xF3 (Fig. 1 below) whose importance was suggested by the Dispersion-Focalization Theory (DFT) of vowel systems (Schwartz et al. 1997; 2012).

Fig. 1 - *Comparison between Unknown Voice (UV) and Known Voice (KV) as performed through Visible Vowels. Above: the vowel space for F1xF2. Below: the vowel space for F2xF3. Vowel labels: "A" /a/, "E" /e/, "O" /o/, "I" /i/, "U" /u/.*



As evident from Fig. 1, although no instances of /i/ were available for UV, it was still possible to appreciate how the vowels were differently distributed, in particular for what concerned the F1-F2 space. In particular, the distribution of KV's formants is extremely well delimited with respect to that of UV (and more concerning F1). The same is true in the F2-F3 region, where the distribution of KV's formants is also extremely well circumscribed in comparison to that of UV for vowels, particularly F3. Both of these graphs provide evidence that contradicts the possibility of UV and KV belonging to the same person.

On a qualitative level, KV was perceived with a more nasal realization of vowels, especially /a/ and /e/. Although this is not strongly reflected in F2/F3 variability, the Harmonics-to-Noise Ratio (HNR) showed a clear difference between UV and KV, as reported in Table 3.

Table 3 - *Harmonics-to-Noise Ratio in Known and Unknown voice*

	Known Voice (KV)	*Unknown Voice (KV)*
/a/	7.017 dB	4.897 dB
/e/	7.336 dB	4.977 dB
/o/	7.524 dB	6.039 dB

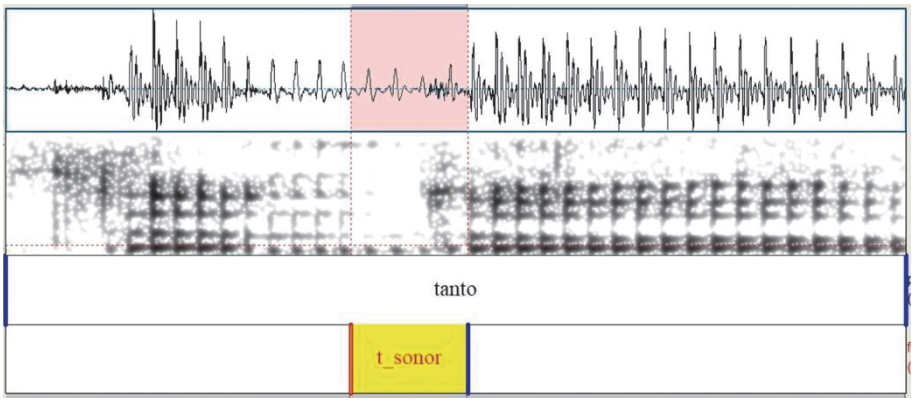
Finally, we presented in the report a sociophonetic profiling of both voices. To avoid biasing the subsequent results, we did this analysis while running the perceptive test and before the pseudo-quantitative formant analysis. Sociophonetic profiling considers the variability of language according to different geographical and social factors, other than the topic and setting of the recordings. Profiling should be accomplished by a forensic audio expert who is also very knowledgeable about the language of the case from a broad linguistic standpoint, taking into account not only phonetics but also dialectology and sociolinguistic variability. This profile operation is typically performed prior to any further perceptual or quantitative testing on the materials, so reducing the expert's potential biases (see also Meluzzi et al., 2020). For this case, giving the aforementioned limitations of the audio materials, the sociophonetic profiling was carried out for both UV and KV on the basis of prosodic cues and the realization of middle vowels /e/-/ɛ/ in stressed position.

As said before, the scarce occurrences of /ɛ/ did not allow for a quantitative comparison, thus we opted for only a perceptive comparison. The two voices were evaluated independently and the findings were compared. UV pronounced [ɛ] with the stressed vowels in the words *proiettili* 'bullets' and sette '*seven*' of the first audio file, and in the words *merda* 'shit' and *questo* 'this' in the second audio file. From the same audios, it was possible to notice a 'dark' realization of the central vowel /a/ when in stressed position like in the words *puttana* 'bitch' and *cazzo* 'fuck'. This was consistent with the further formant analysis. The prosody corresponds to a Northern Italian variety (Hack, 2012; Cardinaletti & Repetti, 2008), with stereotypical Milanese characteristics (Montreuil, 1991; Cerruti, 2011; Kramer, 2009). Furthermore, based on the message of the intercepted audio, it was possible

to detect that, despite the low register, the speech is articulated in subordinated clues, with no mispronunciations or disfluencies, and no indication of foreigner accent. To sum up, UV could be profiled as an adult male between 35 and 40 years old, with a middle to high level of education; more importantly, the speaker's origins could be placed in Lombardy, especially in the Milanese area.

The profile of the KV was more complicated. The expert should have no prior knowledge of the subject being profiled, but even in this scenario, some preliminary information could skew the study. As a result, the profile must be based solely on objective qualities of the voice as revealed by the audio at hand. Since KV has repeated the same sentences of the intercepted audio containing the UV, it was possible to compare the exact same words across the two recordings. The results showed that KV realized the stressed anterior middle-vowel in closed syllables as [e] (e.g., in the word *sette* 'seven'), whereas the vowel /a/ was realized as more fronted than UV. Furthermore, in a voice note, KV voiced the alveolar consonant following the nasal in the word *tanto* 'a lot,' producing ['tan.do], as evidenced by the spectrogram (cf. Fig. 2).

Fig. 2 - *The spectrogram of KV's vocal note showing the sonorization of the alveolar consonant in the word* tanto *'a lot'*



These phonetic features point towards a Southern Italian regional pattern. These characteristics were also confirmed by the intonation patterns shown by KV in the spontaneous speech recorded in the phone call. However, since these features were not extremely marked in KV's speech, the hypothesis was made that KV came from Southern Italy, and, hypothetically, from Sicily, and that he maintained relationships in this area, thus justifying the presence of some phonetic and phonological pattern of southern regional varieties of Italian. KV was assumed to be an adult male between the ages of 30 and 40 based on more impressionistic qualities of his voice quality. It was worth noting that the sociophonetic profile found a match in the comments lately made by both Italian experts and naïve speakers on the two voices during the perceptive experiment. This indicates the perceptual importance of even minor phonetic characteristics in conveying sociolinguistic meaning. Furthermore,

the lawyer later confirmed, after the reception of the report, that her client (i.e., KV) was 33 years old and lived in the Milanese area, but he was originally from the area of Agrigento (Sicily), where he regularly spent his summers.

4. *Results and discussion*

Results can be summarized as follows: Mr. G.'s surname was not pronounced in the first environmental recording, and UV and KV are not the same speaker.

It is clear that such a categorical result is extremely rare. It should be noted that the information provided to us after the report was completed supported our view. Indeed the lawyer informed us that the attribution of the vocal timbre to Mr. G. had been conducted on the basis of the surname heard pronounced by another guy in the initial tape. Without calling into question the good faith and investigative preparation of those who carried out the work, we would like to emphasize how this can thus be defined as a classic case of altered perception: the speaker was most likely inferred from a clue previously extrapolated without scientific validation, e.g. "since Mr. G. has been previously cited, he must be the person speaking on the other recording". Furthermore, our understanding of the technical and operational conditions under which many agents operate substantiates the work. Many consecutive hours with headphones, without being able to equalize the audio or intervene on the signal to improve its intelligibility further affects the speaker's understanding and attribution to reinforce the concept that knowledge of the facts is not alone a guarantee of infallibility in this field (Fraser, 2003; French, Fraser, 2018), especially when data are so few and poor from the qualitative point of view.

William Labov once said that doing linguistic research is often a matter of making the best possible use of poor data (Labov, 1972). This aphorism was written with historical sociolinguistics in mind, but it can be applied to any situation in which a lack of data and the objective impossibility of collecting new ones forces the researcher to adapt his paradigms and standards in order to always provide reliable and informative research. In audio forensics, the data at disposal for a phonetic comparison are usually very 'poor' both in quantitative and qualitative terms. As seen in this work case, it is usual that speech data belonging to the UV are very scarce, whereas it is possible to collect many more audio samples from the KV. Furthermore, the quality of the audio, especially in case of UV, is usually very scarce too, in terms of background noise and spectral information. It is thus beneficial to think in a non-mechanical manner, drawing from all available clues to generate conclusions that can be valuable in the forensic profession. Indeed, deductions and logical consequences might lead to perfect conditions for addressing a forensic inquiry, as in this example.

5. *Conclusions and further perspectives*

The reason it is necessary to make public work like this is obviously mainly an ethical principle. Justice and trust in it are undermined by innumerable factors widely discussed in the audio forensics and applied forensic linguistics literature (e.g., Maher, 2009; Fraser, 2021; Cenceschi, Meluzzi & Trivilini, 2021). Furthermore, experience in real contexts suggests that highlighting problems and viable solutions with concrete examples of real case works is the best way to fuel a healthy collaboration between the scientific and the judicial worlds, even when the difficulties venture beyond the technical themes, into the complicated sphere of bureaucracy. A second, but no less significant, motivation for this publication is the necessity to warn researchers interested in forensic competence that, in judicial contexts, it is often necessary to virtually forget one's scientific language in order to speak effectively in the cause of truth and justice. It is also necessary, for example, to be able to give up the possibility of completing an analysis, and to build deductions 'out of the box' on the basis of a small number of the available elements, without forgetting the scientific nature of the work.

From a scientific point of view, the forensic data is, as mentioned, unpredictable and often of low quality and duration (Meluzzi, Cenceschi & Trivilini, 2020). This case is the demonstration that these characteristics do not completely wipe out the possibilities for intervention and the work can always lead to useful conclusions. It all relies on who, how, and even when the analysis are performed. The expert can intervene at different moments of the judicial process, and often, due to the structure of the judicial process and the bureaucracy, there is no guarantee that the expertise will be, regardless of its goodness, exploited or considered. However, the expert has neither duties nor powers in this sense, his role is to carry out the analysis and express the objective results in the best feasible way. Again, making scientific terminology comprehensible is critical in order to avoid stalling the justice procedure, especially when examining documents already examined by "ad hoc experts" (French & Fraser, 2018).

Bibliography

Boersma, P. (2001). Praat, a system for doing phonetics by computer. In *Glot International*, 5(9/10), 341-345.

Cardinaletti, A. & Repetti, L. (2008). The phonology and syntax of preverbal and postverbal subject clitics in northern Italian dialects. In *Linguistic inquiry*, *39*(4), 523-563.

Celata C., Meluzzi C. & Ricci I. (2016) The sociophonetics of rhotic variation in Sicilian dialects and Sicilian Italian: corpus, methodology and first results. In *Loquens*, 3(1), e025. DOI: http://dx.doi.org/10.3989/loquens.2016.025

Cenceschi, S., Meluzzi, C. & Nese N. (2020). Speaker's identification across recording modalities: a preliminary phonetic experiment. In Romito, L. (ed.) *Language change under contact conditions*, Studi AISV 7, Milano: Officinaventuno, 409-428. DOI: 10.17469/O2107AISV000019

Cenceschi, S., Meluzzi, C. & Trivilini, A. (2021). Audio compression and speaker's discrimination: perspectives for forensic phonetics in the Italian setting. In *Indagatio Didactica*, 13(5), 143-154.

Cerruti, M. (2011). Regional varieties of Italian in the linguistic repertoire. In *International Journal of the Sociology of Language,* 210, 9–28.

Di Paolo, M. & Yaeger-Dror, M. (2011). *Sociophonetics. A Student's Guide*. London: Routledge.

Fraser, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. In *Forensic Linguistics*, 10, 203-226.

Fraser, H. (2021). Forensic Transcription: Legal and scientific perspectives. In Bernardasci C., Dipino D., Garassino D., Negrinelli S., Pellegrino E., Schmid S. (eds.) *Speaker individuality in phonetics and speech sciences: speech technology and forensic applications*, Studi AISV 8, Milano: Officinaventuno, 19-32.

French, P. & Fraser, H. (2018). Why "Ad Hoc Experts" should not Provide Transcripts of Indistinct Audio, and a Better Approach. In *Criminal Law Journal*, 42, 298-302.

Hack, F.M. (2012). *The syntax and prosody of interrogatives: Evidence from varieties spoken in northern Italy*. Oxford University (UK): Doctoral dissertation.

Heeringa, W. & Van de Velde, H. (2017). Visible Vowels: A Tool for the Visualization of Vowel Variation. In *Proceedings CLARIN Annual Conference*, Clarin Eric, 4034-4035.

Kersta, L.G. (1962). Voiceprint identification. In *The Journal of the Acoustical Society of America*, 34(5), 725-725.

Kramer, M. (2009). *The phonology of Italian*. Oxford: Oxford University Press.

Labov, W. (1972). *Sociolinguistic patterns*, Pennsylvania: University of Pennsylvania press.

Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. In *Behavior Research Methods, Instruments, & Computers*, 41(3), 841-849. doi:10.3758/BRM.41.3.841.

Maher, R.C. (2009). Audio forensic examination. In *IEEE Signal Processing Magazine*, 26(2), 84-94.

Meluzzi, C. (2022) La ricerca sul campo (e in campo). In Meluzzi C., & Nese N. (eds.) Metodi e prospettive della ricerca linguistica. Milano: Ledizioni, 37-52.

Meluzzi, C., Cenceschi, S. & Trivilini, A. (2020). Data in forensic phonetics from theory to practice. In *Teanga*, 27, 65-78.

Montreuil, J.P. (1991). Length in Milanese. In *New analyses in Romance linguistics*, 37-47.

Savy, R. (2005). Specifiche per la trascrizione ortografica annotata dei testi, in Albano Leoni, F. (a cura di) *Italiano Parlato. Analisi di un dialogo*. Napoli: Liguori editore.

Schwartz, J.L., Boë, L.J., Vallée, N. & Abry, C. (1997). The dispersion-focalization theory of vowel systems. In *Journal of phonetics*, 25(3), 255-286.

Schwartz, J.L., Boë, L.J., Badin, P. & Sawallis, T.R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. In *Journal of Phonetics*, *40*(1), 20-36.