tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021
https://doi.org/10.34624/id.v13i5.27181

ISSN: 1647-3582

Audio compression and speaker's discrimination: perspectives for forensic phonetics in the Italian setting

Compressione del segnale audio e discriminazione del parlatore: prospettive per la fonetica forense nel panorama italiano

Sonia Cenceschi
University of Applied Sciences of Southern Switzerland (SUPSI)
sonia.cenceschi@supsi.ch

Chiara Meluzzi
University of Milan
chiara.meluzzi@unimi.it

Alessandro Trivilini
University of Applied Sciences of Southern Switzerland (SUPSI)
alessandro.trivilini@supsi.ch

Abstract

In forensic phonetics, speaker's recognition is considered as a conventional chore. The purpose of this work is to analyse whether and to what extent (1) the expertise of the evaluators and (2) reading and spontaneous speaking styles influence the speakers' identification. Our analysis is founded on two different perception experiments. The first one is a real case we worked on in which we challenged both speaker with experience in the audio field and lay speakers to compare two voices in short and low-quality audio files, obtaining very weak result. From these findings, we settled a second 'laboratory' experiment with made-up files recorded by 1 Italian female speaker in 3 settings both for reading and spontaneous speech: high quality recordings, WhatsApp audio and phone call recordings. These data were used in a perceptive test where respondents were asked to point out the recording modality of each sample and specify whether the speaker was the same. Results of this second test show that self-declared experts in audio analysis or transcriptions behave similarly to lay speakers, and that the comparison is more reliable for spontaneous speech when the audio quality is not the same. This result confirms the need to adequately train professionals combining subjective listening with in-depth acoustic and linguistic analyses, and take speech style into account when recording speech samples for comparative analyses.

Keywords: Forensic Phonetics; Speaker's Identification; Speech Perception; Vowel formants; WhatsApp audio analysis.

Resumo

Na fonética forense, o reconhecimento do falante é considerado uma investigação convencional. O objetivo deste artigo é analisar se, e em que medida, (1) a experiência dos avaliadores e (2) os

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

estilos de leitura e de conversa espontânea influenciam a identificação dos falantes. A nossa análise é baseada em duas experiências de perceção diferentes. A primeira trata-se de um caso real em que as autoras desafiaram falantes com e sem experiência na área de áudio com o objetivo de comparar as duas vozes em arquivos de áudio curtos e de baixa qualidade, tendo obtido resultados muito fracos. A partir dos resultados da primeira experiência, as autoras realizaram uma segunda experiência em "laboratório", tendo feito gravações da voz de uma falante de italiano em três modos diferentes - alta qualidade, áudio via WhatsApp e gravação de telefonemas, tanto para leitura, como para discurso espontâneo. Os dados obtidos foram usados num teste de perceção no qual foi solicitado aos respondentes para indicarem como cada amostra foi registada e especificarem se o falante era o mesmo. Os resultados desse segundo teste mostram que especialistas autodeclarados em análise ou transcrição de áudio se comportam de forma semelhante a ouvintes não profissionais e que a comparação é mais fiável para a conversa espontânea quando a qualidade das amostras é diferente. Este resultado confirma a necessidade de treinar adequadamente os profissionais, combinando a audição subjetiva com análises acústicas e linguísticas mais aprofundadas e tendo em consideração o estilo do discurso na gravação de ensaios de fala para efetuar análises comparativas.

Palavras-chave: Fonética forense; Identificação do locutor; percepção da fala; formantes vocálicos; Análise de áudio do WhatsApp.

Abstract

Nella fonetica forense, il riconoscimento del parlatore è considerata un'indagine di routine. Lo scopo di questo lavoro è analizzare se, e in che misura, (1) l'esperienza dei valutatori e (2) gli stili di lettura e di parlato spontaneo influenzano l'identificazione dei parlanti. La nostra analisi si basa su due diversi esperimenti percettivi. Il primo è un caso reale per cui è stato chiesto ad ascoltatori con e senza esperienza nel settore dell'acustica forense di confrontare due voci in file audio brevi e di bassa qualità, ottenendo risultati ambiguamente interpretabili. Abbiamo quindi svolto un secondo esperimento in laboratorio registrando una parlatrice italiana con tre diverse modalità (alta qualità, WhatsApp, registrazione della telefonata) sia per la lettura che per il parlato spontaneo. Questi dati sono stati utilizzati in un test percettivo in cui agli intervistati è stato chiesto di indicare la modalità di registrazione di ciascun campione e di specificare se il parlante fosse lo stesso. I risultati di questo secondo test mostrano che coloro che si erano auto dichiarati esperti in analisi o trascrizioni audio si comportano in modo simile agli ascoltatori non professionisti e che il confronto è più affidabile per il parlato spontaneo quando la qualità dei campioni è diversa. Questo risultato conferma la necessità di formare adeguatamente i professionisti combinando l'ascolto soggettivo con approfondite analisi acustiche e linguistiche e di tenere conto dello stile del parlato durante la registrazione di saggi fonici per analisi comparative.

Parole chiave: Fonetica forense; identificazione del locutore; percezione linguistica; formanti vocaliche; WhatsApp audio analysis.

Introduction

This paper addresses the specific and problematic reality of speaker's recognition for forensic purposes (cf. Kumar 2019) with respect to the Italian setting. As a matter of fact, specific language or audio experts are not designated in the department of forensics in Italy and, as a result, reports

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

used in courtroom are often provided by lay experts, with little or no competence in phonetics and/or sound engineering (Bartle & Dellwo 2015, but also Jessen 2010 on the figure of forensic phonetician).

It is evident that this practice could become really risky, especially when transcriptions or perceptive evaluations on speakers' identity are based on low-quality and/or noisy audio samples (Fraser 2003, 2015). Therefore, in order to emphasize the grievous practice of using lay experts, and the need to carry out technical analyses to validate the acoustic perception when working with 'corrupted' audio files, we designed two specific laboratory experiments. In particular, we aim at answering the following research questions:

1. Do lay speakers perform differently than audio or linguistic experts in speakers' discrimination tasks with low-quality or noisy audio data?
2. Do speakers perform differently depending on whether the modality is reading or spontaneous speech?

The paper is organized into five further sections. The next section briefly introduces the theoretical premises about speech perception, speakers' discrimination, and noisy audio evaluation. Section 3 presents the materials and methods of the two experiments we propose, whereas section 4 represents the results of said experiments. In section 5, we discuss the results before moving to some conclusions and further perspectives in section 6.

Theoretical premises

Speakers' identification and discrimination is a fundamental task in forensic sciences, and it could be accomplished through different methodologies both referring to the acoustic cues of the speech signal and the listener's perception. Although these methodologies should be assimilated (cf. Sigona & Grimaldi 2017), it is often the case that speech samples are too short, noisy and/or of low quality to prevent a semi-automatic acoustic analysis. Therefore, reports on speakers' discrimination often rely on the listeners' judgement only, and this practice seems to be accepted in different national legal systems (see Lindsay et al. 2000, Edmond et al. 2015). This introduces the dilemma of the listeners and, more generally, on the level of expertise of the ones providing the reports. As pointed out by French & Fraser (2018: 299), it is still quite frequent to rely on 'ad hoc experts', whose level of expertise is solely derived from listening to a recording many times.

However, when working in forensics, often the audio evaluation is based on noisy speech samples, but also on materials obtained with different technologies and recorded with different compression formats. Psychoacoustic research has demonstrated how noise and, more generally speaking, low audio quality, could affect the spectrographic information of the audio file. For instance, Li et al. (2012) demonstrated that noise affects the fundamental frequency (F0) and inhibits the distinction between English fricatives, in particular as it concerns voiced vs. voiceless phones, and also among places of articulations. Noisy or indistinct audio could be referred to both 'hard to hear' content, and also to misperception of the content, because of a mismatch between the acoustic priming and the contextual expectations (Fraser 2015, 2018).

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

Previous works in both psychoacoustic and sociology of language have demonstrated the importance of stereotypes in shaping our perception, and the same conclusions have been reached by phoneticians working in forensics (e.g., Lavan et al. 2019a, *ex multis*). Indeed, it should be emphasized how "perception is an active, rather than a passive process, with the hearer actively constructing, rather than passively picking up the speaker's message" (Fraser 2003: 205). Recent work of Rose (2002) on forensic speakers' identification identified the different biases operating in the listener's mind when confronting two audio samples for forensic purposes, such as the identification of the talker among a list of suspects or the matching between a known and an unknown voice. Furthermore, Smith et al. (2018) have demonstrated how lay listeners, that is listeners without a full training in speech sciences, are negatively affected by background noise and also by speech type while performing a speaker's recognition task. In a recent article, Lavan et al. (2019b) have also emphasized how speech variability because of stylistic or emotional variation impair generalization across voice samples, and, thus, prevent identification.

Materials and methods

Mr. G's case

The considerations presented in this paper started from a real case the authors have worked on, and that we referred to as Mr. G's forensic case, in order to maintain the client's privacy. Mr. G was accused of a crime based on a low-quality noisy audio file of only five seconds: according to the prosecutor, the audio contained the client's voice. However, this recognition was made during the investigation phase, and it was reported that the person who did the recognition listened to the audio only once. Therefore, that was a subjective listening based on very poor material. Indeed, it is important to underline that the audio material relating to the unknown voice had a too limited duration to perform a quantity vocal comparison, whether deriving from perceptual or semi-automatic analysis.

Mr. G's attorney asked for the opinion of external experts regarding the audios and the evaluation performed by the judge's technical adviser. We preliminarily deemed invalid the evaluation carried out by a single person on the basis of such noisy materials and on a single listening. After that, the attorney asked for a report to prove that the original identification was wrong or invalid. This is a typical working scenario in case of speakers' comparison: meagre audio material, recorded by different technical means could be used to question the reliability of the identification above reasonable doubt. In our case, the comparison needed to be made between the low-quality audio of an unknown voice (henceforth, UV), and the voice of the client as recorded by the attorney in a high-quality format, and which we will refer to as the known voice (henceforth, KV). Given the impossibility of carrying out a quantitative comparison, it was necessary to record KV in low-quality to limit variations. Moreover, the comparison had been made precisely between KV's high quality audio and UV's low quality by a Law Enforcement Officer, and so the initial conditions were replicated.

After a spectrographic inspection of the UV, we decided also to confute the identification not only through an acoustic analysis (vowel formants' comparison, sociophonetic profiling, etc.)

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

but also through a perceptive test with different subjects. We selected 3 groups of respondents, 2 trained phoneticians who were Italian native speakers, 6 audio specialists who were non-Italian speakers and 117 lay speakers with Italian as their L1, aged between 19 and 24.

Each group listened to 2 audio stimuli only once, and for each one they had to immediately answer to the simple question “Is it the same speaker?”; the answer could only be YES/NO, but respondents have the possibility to explain their choice. They could listen to the stimuli only once, because our intent was to show how the procedure adopted by the prosecutor (i.e., an identification based on a single listening of poorly recorded materials) was fallacious and potentially dangerous. Of the two stimuli, stimulus A was, in fact, a distractor since it contained the UV compared to a similar voice extracted from another recording of the same investigation (belonging to a third known speaker), and henceforth, referred to as KVD (Known Voice Distractor). Stimulus B presented the real comparison between KV and UV. In both stimuli the two voices were separated by 2” of silence.

Laboratory test

Based on the results of the first experiment, we decided to prepare a second experiment in a more laboratory-like setting in order to investigate how perception is influenced by the style of speech and the listener’s experience in the audio field, eliminating the noise factor typical of environmental interception.

For this experiment we recorded 1 female with a North-Western Italian L1 accent who was 20 years old. She performed a sentence reading task with neutral statements, balanced by prosodic contour and with 12 target disyllabic real words with the cardinal vowels /a, i, u/ in either singleton or geminate context (see also Cenceschi et al. to appear). The target speaker also performed a more spontaneous recording in which she described her room and her favourite dish. Both controlled and spontaneous speech was recorded in 3 modalities including high quality recordings (MIC) through a TASCAM DR20 with an external microphone PDM 663 (44.1 kHz sample frequency, .wav), a WhatsApp audio message (WA) sent to the first author, and, finally, a phone call made to the first researcher’s mobile and recorded through the app Voice Recorder (henceforth, VR). Although the app allows high quality recordings in .wav format, it was decided to use the default settings (16 kHz sample frequency, .mp3) to mimic an intercepted audio recorded for forensic purposes (see also the following discussion of the present results).

The perceptive test showed the same voice as recorded in different modalities (WA vs. VR, WA vs. MIC, MIC vs. WA, and the other way round). We asked again a different subset of participants to say whether they believed the speaker to be the same or a different one and to justify each answer. The questionnaire was created and administered via Google Forms, collecting the main profiling data in an anonymous form, and asking participants to sign the releases regarding aggregate use of data for scientific purposes. The link was shared on the main social networks and to selected groups of expert listeners (musicians, transcribers, phoneticians, etc.), with no particular requests regarding the use of headphones. This choice was weighted to reach a greater number of listeners but still reflecting the Italian forensic situation, in which often the auditory attribution of the speaker’s voice is performed (especially in the preliminary investigation phase) without particular attention to the choice of headphones

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

suitable for speech listening. Indeed, the case of Mr G. reflected exactly this process: the auditory comparison carried out by the Court expert had not been documented (being most likely performed without headphones or with low quality headphones), underlining a lack of attention to psychoacoustic and training issues, as well as excessive confidence in the hearing abilities of the individual operator.

We collected 426 answers from Italian L1 speakers, aged between 18 and 68 years old, and all from North-West areas of Italy, therefore matching the geographical variety of the voice samples. We also asked our respondents to self-rank their familiarity with WhatsApp audio messages, from 1 (no familiar at all) to 5 (highly familiar). Among the respondents, 247 declared a high familiarity with WhatsApp audio, whereas only 5% of them (21 subjects) did not know anything about this App. Finally, respondents had to declare whether they had received any professional training in audio manipulation or in speech sciences, in particular phonetics and phonology. These answers were later catalogued by the researchers by dividing respondent into four groups based on the level of respondents' experience with speech and audio file manipulation. Among our respondents, only 16.4% (70 subjects) had high experience with audio management for work or hobby, and, among them, 23 were phoneticians or linguists specialized in audio transcription and annotation (without specific forensic experience).

Results

Results of Mr G.'s experiment

Both trained phoneticians and audio experts agreed in recognizing that in both stimuli the two speakers were not identical. Independently of the language, the experts in audio management did not have doubts in identifying two different speakers by basing on both prosodic and segmental features. Lay listeners, however, behaved differently with respect to the stimulus, as demonstrated in Fig. 1.

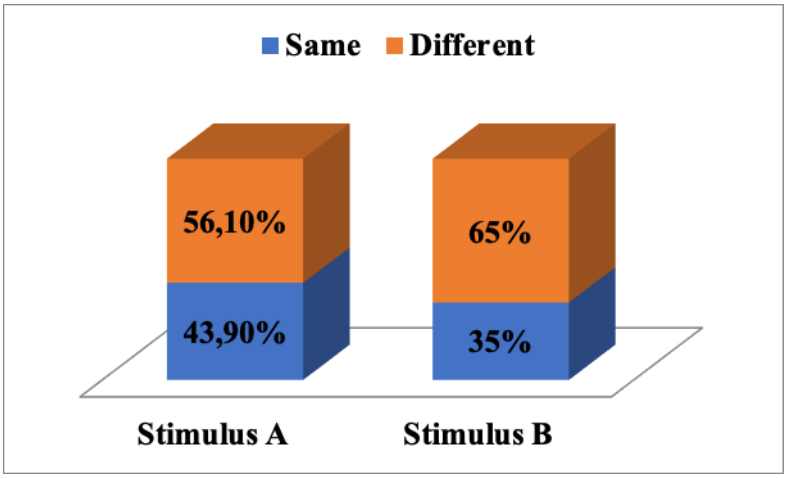

Fig. 1 - Speakers' identification in the two stimuli by lay speakers (117 answers); stimulus A is the distractor, while Stimulus B the real sample.

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

Data show that in both cases, lay speakers identify the two voices as belonging to two different speakers, with a percentage of accuracy over 50% in both the fake and the real stimulus. However, the results are more straightforward for stimulus B (the true comparison) rather than for stimulus A (the distractor with KVD): indeed, in the made-up stimulus A, lay speakers identify two different persons in the audio samples in 56.1% of cases, whereas in the real case scenario (B) the percentage increases to 65%.

Results of the laboratory test

A first analysis on the whole sample (426 respondents) is shown in Fig. 2. It is possible to notice that, generally speaking, our respondents were able to correctly identify the same voice across recording modalities in the two speaking styles. However, the percentage increases in spontaneous speech with respect to the sentence-reading task (83.9% and 68.1%, respectively).

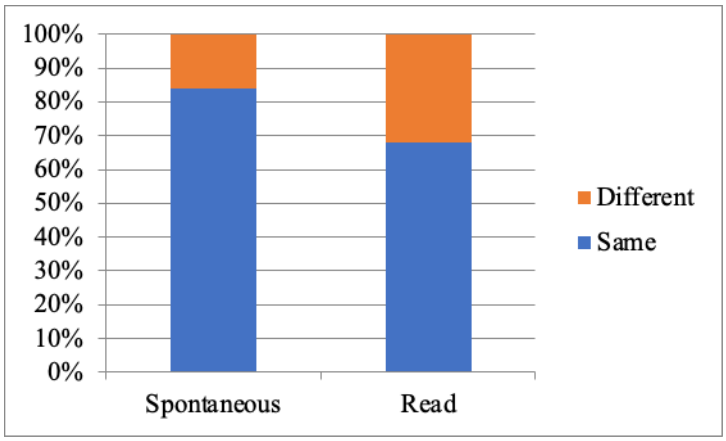

Fig. 2 - Speakers' identification across recording modalities in spontaneous and read speech (426 respondents).

We also tested if respondents' self-declared familiarity with WhatsApp audio messages or with audio analysis modified their perception of voice similarities across recording modalities. Respondents rated their familiarity with WhatsApp audio messages from 1 (not familiar at all) to 5 (very familiar). Tab. 1 shows the correlation between the levels of familiarity and speakers' identification for both spontaneous and read speech samples. It appears that respondents at the two extremes of the scale (1 and 5) behave similarly, with the highest percentages of correct identification of the two voices as belonging to the same speaker. Indeed, for spontaneous speech samples, speakers with no familiarity with WhatsApp (rate 1) recognized the same speaker in the 71.4% of the cases, and this percentage raises to 85.7% in the case of read speech. Similar percentages are found for speakers with a very high familiarity with WhatsApp (rate 5), with 72.4% and 85.8% of correct

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

identification as the same speaker in spontaneous and read speech, respectively. Conversely, users with an average familiarity with WhatsApp appeared to be more uncertain and tended to attribute the two audio samples to different speakers, with variable percentages around 38.2% for spontaneous speech and 20.4% for read speech.

Tab. 1 - Speakers' identification and level of respondents' familiarity with WhatsApp audio in spontaneous and read speech.

	Spontaneous Speech		Read Speech	
	Same	**Different**	**Same**	**Different**
1	71,4%	28,6%	85,7%	14,3%
2	61,8%	38,2%	87,3%	12,7%
3	68,0%	32,0%	84,0%	16,0%
4	65,5%	35,4%	79,6%	20,4%
5	72,4%	27,6%	85,8%	14,2%

A similar picture emerges as it concerns speakers' identification and respondents' level of expertise in managing audio files, within a 4-points rank, as reported in Tab. 2. As it happened before (Tab. 1), respondents generally correctly identified the identity of the speaker across recording modalities, with percentages of correct identification over 50%, and with better results in spontaneous speech rather than in read one.

Indeed, the samples extracted from the sentence reading task appear to confound the hearer, since for all level of expertise more than 30% of our respondents misidentified the samples as belonging to two different speakers. There are, however, huge differences in the distribution of these answers, with again a main difference between respondents with medium experience, and the two poles of the continuum, that is respondents with high or no expertise in managing audio files.

In fact, from data reported in Tab. 2 one should note how the respondents with a medium experience are the ones more prone to misinterpretations. The users who declared a medium competence with audio management (e.g., musicians or non-professional video editors and creators) erroneously judged the same voice as pertaining to two different speakers in the 19.1% and 40.4%, for spontaneous and read speech, respectively. Speakers with a high level of expertise (23 respondents) with audios and transcriptions performed the same across speech style, with a 30.4% (7 answers) of false judgement of non-identity in both cases.

Tab. 2 - Speakers' identification and level of respondents' expertise with audio files.

	Spontaneous Speech		Read Speech	
	Same	**Different**	**Same**	**Different**
High	69,6%	30,4%	69,6%	30,4%
Medium	80,9%	19,1%	59,6%	40,4%
Little	92,3%	7,7%	61,5%	38,5%
None	85,0%	15,0%	69,4%	30,6%

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

Discussion

Mr G.'s experiment

It is very indicative that stimulus A gives a response close to 50% for both cases. This indicates how two different voices (a priori) but recorded and heard with low audio quality, could easily be confused even when heard by a large sample of people. This result should be considered as an alarm for those forensic cases in which the listening is performed by only one person.

It is important to emphasize, however, that percentages of correct non-identification were quite low, and misjudgements quite frequent in both cases. These results indicate that judgements based on different quality audio samples and on a single listening could be misleading. Indeed, the second experiment confirms these findings.

Laboratory test

A better result in speakers' recognition for spontaneous speech was expected, as it carries suprasegmental features more characteristic of the speaker, whereas sentence-list reading presents quite a flat prosodic contour. Despite this, it is necessary to consider the decline in reliability on the spoken speech as the communicative modality is not always spontaneous in forensic recordings, thinking for example of the suspected' spoken speech collected by some lay experts to perform speakers' comparison. In light of this finding, it would always be advisable to collect spontaneous speech data and not (or, not only) elicited or read speech.

This apparently unexpected behaviour of medium experienced listeners is actually useful in the Italian forensic practice, where many lay experts come from the music field and their reports are often used in courts (Romito & Galatà 2007). From the results of this experiment, though, it can be inferred that excessive confidence leads to underestimating a lack of knowledge in the field of speech analysis and degraded signal.

General remarks

Our second experiment confirmed that the supposed expertise with an application or audio analysis could be a double-edged sword. Our results show that self-declared experts in audio analysis or transcriptions behave similarly to lay speakers, with no previous experience in phonetics, audio engineering or related fields. This should not be interpreted in favour of hiring lay speakers against experts for performing forensic reports.

In our experiment, respondents could perform only one listening, and experts are fully aware that this is not sufficient to provide a reliable forensic report (cf. Fraser 2003, but also Kaplan 2020: 78-79), and that perception must match with acoustic analysis of speech sound, especially in

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

case of low-quality or noisy audio samples (Smith et al. 2018). The performances of the so-called experts in our second experiment may be seen supporting this interpretation.

Indeed, these respondents showed the highest percentages of misinterpretation of speakers' identity across recording modalities. This could imply that the general knowledge of sounds, as provided, for instance, by a musician, does not guarantee a more accurate identification with a single listening. Since this belief is still present and operating in Italian forensic taskforce, we believe that these results should warn against on the dodgy habit of performing speakers' identifications based on a single listening to low quality audio samples.

Moreover, it appeared from our second experiment that speech style is also an important variable in performing speakers' identification across recording modalities, thus confirming what was previously reported for other languages also for the Italian language (e.g., Smith et al. 2018, Lavan et al. 2019). Indeed, results reported in 4.2 suggest that low quality audio could emphasize differences in hyper-articulated speech resulting from a formal sentence-reading task (see also Whalen et al. 2004 for the effects of hyper-articulation on speech perception). This is important since the problem of style variability has only been recently addressed in forensic phonetics, in particular, in the Italian setting (see also Paoloni 2015), but unfortunately it is a widespread practice to confront intercepted audio samples with the recording of a suspect reading a declaration. Audio materials with different speech style are often provided as the only evidence, if not indeed realized by ad hoc experts (Fraser 2015), who record the suspect with speech style other than the unknown voice in order to perform a comparison (e.g., asking to read a list of words, numbers, or phrases).

Conclusions and further perspectives

In this paper, we have presented two perceptual experiments based on a real case scenario and on a laboratory one, both dealing with speakers' identification across different recording modalities.

Self-declared experts in audio analysis or transcriptions behave similarly to lay speakers, with the aggravating circumstance of being (in the Italian context) potentially recognized as forensic experts by judges, law enforcement agencies and lawyers. The common thought leads in fact to think that they, even if misled by low-quality files, are generally more able to unconsciously balance the lack of acoustic cues due to their experience. The present study emphasizes the opposite concept, already addressed in Paoloni (2015), and Romito & Galatà (2007), that excessive confidence in one's own abilities can lead, if not to a worsening of the acoustic performance, to an inconsistent difference compared to lay speakers. Once again it is the specific knowledge concerning the characteristics of the speech and the quality of the forensic recordings that make a comparative assessment realistic and reliable.

From a general perspective, these results confirm that a psychoacoustic investigation is not enough to provide the most reliable results in forensic evaluations, though it is unfortunately still a common practice in Italian taskforce[1].

[1] "Le competenze del Perito Fonico. Proposta di Linee Guida dell'Associazione Italiana di Scienze della Voce" (The competences of the phonetic forensic expert. A proposal of guidelines provided by the Italian Association of Voice Sciences), https://www.aisv.it/it/documenti/doc_download/104-propostaolfaisv-perizieforensi.html

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

Based on our findings and in line with previous research, the authors felt the need to strongly recommend the necessity of trained audio experts to perform speakers' identification for forensic purposes, especially in dealing with noisy and/or low-quality audio files. Furthermore, due to the actual lack of parameters for selecting said experts in the Italian forensic settings, it is important to raise awareness on the dangers of evaluations provided by lay listeners on corrupted audios, and on the possible biases induced in perception by noise.

Therefore, it becomes even more necessary to perform engineering or linguistic technical analyses (Paoloni 2015) in order to confirm the auditory perception, or to highlight discrepancies that may underline reasonable doubts about the listening outcome. Moreover, it is confirmed that listening made by a single person, even if expert in the analysis of corrupted audio, does not represent a valid element for attributing a crime to a suspect.

The recent reform of Italian laws about interceptions and the use in courts of intercepted audio files (Legislative Decree n.161, 2019, converted with amendments by law nr. 7, 1 March 2020, entered into force on 1 September 2020) still leaves the transcription phase uncovered by regulation. In particular, this new law does not address the difficulties in the transition from spoken to written language, and the possible misinterpretations and mistakes emerging from analysis on noisy recordings but also among audio with different digital compressions. It should be noted that these problems still affect a large number of cases of incorrect attribution of the speaker. Therefore, the exploitation of forensic audio evidence in Italy is still problematic, and this article aims to help to highlight part of these shortcomings by providing concrete results and solutions from real experiments.

References

Bartle, A., & Dellwo, V. (2015). Auditory speaker discrimination by forensic phoneticians and naive listeners in voiced and whispered speech. *International Journal of Speech, Language & the Law, 22*(2).

Cenceschi, S.; Meluzzi, C. & Nese, N. (accepted – to appear) Speakers' identification across recording modalities: a preliminary phonetic experiment, *Studi AISV 7*, Milano: Officinaventuno.

Edmond, G., Martire, K., & San Roque, M. S. (2011). Mere guesswork: Cross-lingual voice comparisons and the jury. *Sydney Law Review, 33*, 395–425.

Fraser, H. (2003) Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases, *International Journal of Speech, Language and Law 10.2*: 203-226.

Fraser, H. (2015) Transcription of Indistinct Covert Recordings Used as Evidence in Criminal Trials, in H. Selby & I. Freckelton (eds.) *Expert Evidens, Thomson Reuters.*

Fraser, H. (2018) 'Assisting' listeners to hear words that aren't there: dangers in using police transcripts of indistinct covert recordings, Australian Journal of Forensic Sciences, 50:2, 129-139.

French, P., & Fraser, H. (2018). Why 'ad hoc experts' should not provide transcripts of indistinct forensic audio, and a proposal for a better approach. Criminal Law Journal, 42(5): 298-302.

Jessen, M. (2010). The forensic phonetician. The Routledge handbook of forensic linguistics, 378.

Kaplan, J.P. (2020) Linguistics and the Law, London: Routledge.

tecnologias da informação em educação

Indagatio Didactica, vol. 13 (5), dezembro 2021 ISSN: 1647-3582
https://doi.org/10.34624/id.v13i5.27181

Kumar, M. (2019). Forensic Speaker Identification: A Review of Literature and Reflection on Future. Language in India, 19(7).

Lavan, N., Burton, A.M., Scott, S.K. & McGettigan, C. (2019a) Flexible voices: Identity perception from variable vocal signals, Psychonomic Bulletin & Review 26: 90-102.

Lavan, N., Burston, L., Ladwa, P., Merriman, S.E., Knight, S. & McGettigan, C. (2019b) Breaking voice identity perception: Expressive voices are more confusable for listeners, Quarterly Journal of Experimental Psychology, 72(9): 2240-2248.

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. Law and Human Behavior, 24, 685–697.

Paoloni, A. (2015) Sul riconoscimento del parlante in ambito forense, Sicurezza e giustizia 3: 24-37.

Romito, L., & Galatà, V. (2007). Speaker recognition: stato dell'arte in Italia valutazione dei corpora, dei metodi e delle professionalità coinvolte. Scienze vocali e del linguaggio-Metodologie di valutazione e risorse linguistiche (Atti del III Convegno Nazionale AISV-Associazione Italiana di Scienze della Voce, 223-242.Rose, P. (2002). Forensic Speaker Identification. New York: Taylor & Francis.

Sigona, F. & Grimaldi, M. (2017) Tools for forensic speaker recognition, in F. Orletti & L. Mariottini (eds.) Forensic Communication in Theory and Practice, Cambridge: Cambridge Scholars Publishing, pp. 127-148.

Whalen, D. H., Magen, H. S., Pouplier, M., Min Kang, A., & Iskarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. Language and speech, 47(2), 155-174.

Acknowledgments

The paper has been jointly written by the three authors. For the requirements of the Italian Academia, Sonia Cenceschi is responsible for sections 1, 3.1, 4.1 and 6, whereas Chiara Meluzzi is responsible for sections 2, 3.2, 4.2 and 5. Alessandro Trivilini supervised the overall work and revised the final article.