

The scientific disclosure of speech analysis in audio forensics: remarks from a practical application

Cenceschi S., Trivilini A., Denicolà S.

The University of Applied Sciences of Southern Switzerland (SUPSI)

This work describes a university teaching experience of main speech analysis methodologies to non-expert professionals active in the legal and security areas. Results and remarks are being used to develop a final protocol for introductory courses in forensic linguistics. The final purpose is to transmit the potential of the scientific analysis in order to stem the multiple problems caused by misinformation [1] and enhance the dissemination of scientific contents.

The diffusion of scientific (even if constantly evolving) methods is an urgent need, specially in national unregulated audio-forensics contexts. Moreover, the same word “expert”, generically defined as *audio forensics expert*, has countless facets: better speaking therefore about a universe of different specializations which should collaborate, and understand each other, to satisfy ever-changing questions. Since the situation is very complex, this work is narrowed just on one aspect: the optimal simplification of specialized concepts in order to increase the defence capabilities of the whole society from superficial, irresponsible and misleading experts reports and conclusions.

The focus is not on *performing researches* in speech analysis (users already have their specialties), but on *making the person independent and prepared to what he/she must expect from alleged technical experts*.

Italy and Cantone Ticino share the same official language but also similar needs: the main difference lies in usage frequency of speech recordings in trials and investigations. While Italy has been doing a decades-long and huge use of audio recordings, in Italian Switzerland they are, to date, sporadically introduced. Anyhow, the interest is growing up, so that the proposal applies to both countries, whether it be to remedy or to lay foundations for dialog and education.

The protocol is limited to needs and suggestions collected among judges, lawyers and law enforcement officers during a job lasted years and later used to create a training course. Ideas have been so applied during the course “Biometrics and emerging technologies” included in the Certificate of Advanced Studies (CAS) in Advanced Digital Forensics, organized by the Digital Forensics Service of the University of Applied Sciences of Southern Switzerland (SUPSI). Participants possess an average informatics preparation, but no requirements concerning audio analysis and human voice: they were a homogeneous group, good to test the difficulty of concepts which have been carried out in about 15 hours. Even if it was the first attempt, collected feedbacks confirmed the strong interest and the acquisition of critical sense and correct basic knowledge.

The following list reports most common technical forensics requests limited to the human voice and related examples of wrong procedures or conclusions. They are just few examples, but complex enough to extrapolate a common developing basis and show the gap between reality and utopia.

TECHNICAL REQUEST	EXAMPLES OF WRONG CONCLUSIONS/RESULTS
Increase of intelligibility	The audio has been cleaned too much or transcripts are made listen to clips in loop so that speech perception is altered.
Disambiguation of linguistic contents	The phonetics analysis is not done and conclusions rely only on subjective perception.
Voice comparison	Conclusions are carried out without controlling SNR value requirements, or doing them perceptively, or without underlining integrity/conservation issues and providing answers without uncertainty (Y/N).
Geographical/Dialect detection	Linguists and dialectologists are not consulted and the analysis is conducted on erroneous perceptive bases.

The protocol application started showing typical wrong methodological answers to diffuse requests, and applying a reverse process, listing procedural and technical errors. Once introduced realistic answers to forensics queries, the human speech complexity and phonetics fundamentals have been addressed in order to underline discrepancies. Communicate these concepts to non-experts, without speech and sound analysis knowledge, is the most crucial point. The experience

showed that spectrogram concept can be well explained without going into detail. Furthermore, concepts related to noise and harmonicity, vowels and consonants are also easier to be understood if exploiting parallels with music [2]. This work has produced excellent results on the critical sense of listeners, showing it is possible also with a few hours available.

A superficial introduction to complex topics has been enough to allow the analysis of practical cases, where the signal has not been correctly processed. Proceeding then for practical examples, it has been showed conclusions can never be "yes" or "no", but a range of variable nuances according to the starting situation conditions, with a brief excursus about Likelihood Ratio [3][4].

The danger in communicating doubts necessity, and explaining casual, systematic and inevitable process errors, resides in transmit the wrong perception that, then, speech recordings are useless. The found remedy has been then to underscore the potentiality of uncertainties, specially respect to social security and potential judicial errors. Absolute certainty is not a proof, but a false solution emptied of meaning.

After completing this work of "dismounting and reassembling with cognition" process, two main topics have been addressed: voice comparison and speech perception. The desire was to provide a comprehensive overview being there more in vogue on the media compared to many others. The synthesis consisted of automatic, semi-automatic and perceptive analysis for speakers' comparisons, with explanations regarding main problems and obstacles [5]. Perception issues have been listed instead [6][7], to emphasize transcripts are not so trivial, especially to emphasize that the transcripts are not so trivial, even more with background noise.

The last theme dealt with is the need for a division of tasks, describing, with practical examples, the needed knowledge to address different issues. If the constant dialogue between different professions is utopia, it is also true that the situation can always be improved and it can be done thanks to dissemination and involvement. Dialogue must not increase just between police forces, judges, lawyers and the scientific world, but also among experts within the latter.

The complete protocol deeps and orders the themes here summarized, but all the project is a starting point based on one personal experience, consequently representing a partial view and needing deepen discussion and practical applications. Moreover, this topic involves other essential and serious practical issues can just be cited here, such as the standardization and regulation of specialized courses, procedures, and remuneration [8]. The vastness of the topic scares, but it is certainly better to be frightened than not, since judgments or investigations directly affect safety, costs and life of citizens.

References

- [1] Carrada, G. (2005). *Comunicare la scienza: kit di sopravvivenza per ricercatori* (Vol. 12). Alpha Test.
- [2] Wolfe, J. (2002). Speech and music, acoustics and coding, and what music might be 'for'. In *Proc. 7th International Conference on Music Perception and Cognition* (pp. 10-13).
- [3] Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic science international*, 240, 61-68.
- [4] Becker, T., Jessen, M., & Grigoras, C. (2008). Forensic speaker verification using formant features and Gaussian mixture models. In *Ninth Annual Conference of the International Speech Communication Association*.
- [5] Romito, L., & Galatà, V. (2004). Towards a protocol in speaker recognition analysis. *Forensic science international*, 146, S107-S111.
- [6] Romito, L. (2005). La competenza linguistica nelle trascrizioni Forensi: l'intelligibilità, l'oggettività e il rapporto segnale/rumore. *Detective And Crime*.
- [7] Fraser, H. (2003). Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics*, 10, 203-226.
- [8] 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*, 3, 223-242.