Forensic Speech Enhancement: Toward Reliable Handling of Poor-Quality Speech Recordings Used as Evidence in Criminal Trials

Helen Fraser; Vincent Aubanel; Robert C. Maher; Candy Mawalim; Xin Wang; Peter Počta; Emma Keith; Gérard Chollet; Karla Pizzi

OPEN ACCESS Freely available online

H. Fraser, V. Aubanel, R. C. Maher, C. Mawalim, X. Wang, P. Pořta, E. Keith, G. Chollet, and K. Pizzi, "Forensic Speech Enhancement: Toward Reliable Handling of Poor-Quality Speech Recordings Used as Evidence in Criminal Trials," *J. Audio Eng. Soc.*, vol. 72, no. 11, pp. 748–753 (2024 Nov.). https://doi.org/10.17743/jaes.2022.0176.

Forensic Speech Enhancement: Toward Reliable Handling of Poor-Quality Speech Recordings Used as Evidence in Criminal Trials

HELEN FRASER,[1] *AES Associate Member*, **VINCENT AUBANEL,**[1,*] *AES Associate Member*,
(helen.fraser@unimelb.edu.au) (vincent.aubanel@unimelb.edu.au)

ROBERT C. MAHER,[2] *AES Fellow*, **CANDY MAWALIM,**[3] **XIN WANG,**[4] **PETER POCTA,**[5]
(rmaher@montana.edu) (candylim@jaist.ac.jp) (wangxin@nii.ac.jp) (peter.pocta@uniza.sk)

EMMA KEITH,[6] **GÉRARD CHOLLET,**[7] **AND KARLA PIZZI**[8,9]
(emma.keith@anu.edu.au) (gerard.chollet@telecom-sudparis.eu) (karla.pizzi@tum.de)

[1]*Research Hub for Language in Forensic Evidence, The University of Melbourne, Melbourne, Australia*
[2]*Electrical & Computer Engineering, Montana State University, Bozeman, MT*
[3]*Japan Advanced Institute of Science and Technology (JAIST), Nomi, Japan*
[4]*National Institute of Informatics (NII), Tokyo, Japan*
[5]*University of Zilina, Zilina, Slovakia*
[6]*Australian National University, Canberra, Australia*
[7]*Institut Polytechnique de Paris, Paris, France*
[8]*Neodyme AG, Garching, Germany*
[9]*Technical University of Munich (TUM), Munich, Germany*

This paper proposes an innovative interdisciplinary approach to evaluating the effectiveness of forensic speech enhancement (FSE). FSE faces unique challenges arising from a range of factors, from poor recording quality, highly variable conditions from case to case, and content uncertainty. Despite these difficulties, FSE is commonly admitted in court, and can significantly influence the outcome of criminal trials. Current FSE practices are hindered by unrealistic expectations from courts, which often assume that enhanced audio inherently clarifies content. In fact, FSE can have the undesired opposite effect, potentially resulting in unfair prejudice, when, for example, it increases the credibility of a misleading transcript. The proposed interdisciplinary project advocates for a better consideration of speech perception factors, particularly those related to transcription. It aims to bridge the gap between FSE and forensic transcription by promoting a combined approach to enhancing and accurately transcribing forensic audio. By developing a position statement on FSE capabilities, the project seeks to establish realistic standards and foster collaboration among researchers and practitioners. This effort aims to ensure reliable, accountable forensic audio evidence, aligning with forensic science standards and improving the effectiveness of the justice system.

0 INTRODUCTION

This communication seeks to engage interest in an innovative approach to evaluating the effectiveness of forensic speech enhancement (FSE) that is being pioneered by a small interdisciplinary group at the University of Melbourne, Australia. Speech enhancement is the well-established field that aims to make recorded speech easier to understand or pleasanter to listen to. It has attained many impressive achievements across a wide range of applications, including hearing aids, security, entertainment, and speech perception theory [1–5]. FSE is often taken to be simply an applied branch of this larger field. However, FSE is significantly different than, and far harder than, other kinds of speech enhancement.

Forensic audio includes any recording relevant to a criminal investigation, but here, the definition is narrowed to focus specifically on recorded speech used as evidence in court. This is traditionally the product of covert surveillance by law enforcement but now comes from a wide variety of sources, including smartphones and body-worn cameras.

*To whom correspondence should be addressed, email: vincent.aubanel@unimelb.edu.au.

The only characteristic that all forensic speech recordings share is that they are alleged to contain confessions, admissions, or information that the speakers are not willing, or able, to confirm openly in court.

A very common characteristic of forensic audio is that the quality is extremely poor, to the extent that there can be uncertainty or disagreement regarding exactly what words are spoken. For use in court, the content needs to be clear. To achieve this, investigators have naturally turned to the audio-enhancement community, who have been keen to help—and most jurisdictions now routinely admit enhanced versions of forensic audio in court. The problem is that enhancing techniques that work well for other spoken word recordings tend not to be so successful for FSE.

1 WHAT FACTORS MAKE FSE SO MUCH HARDER THAN ENHANCING OTHER SPEECH RECORDINGS?

The most obvious factor is that the recording quality is far worse, meaning the audio is often unintelligible to listeners lacking prior knowledge of the content. This is because forensic scenarios allow little control over the recording conditions: microphones may be poorly placed, equipment settings may be nonoptimal, there may be loud and intermittently varying background noise such as traffic, television or other conversations. Worse still, these issues are highly variable from one case to another, limiting opportunities to standardize processing algorithms or to train machine learning models (although see [5] for a recent promising approach). The situation was well summarized in Gaston Hilkhuysen's tutorial on speech intelligibility at the 2014 AES Conference [6, "Tutorial 4: Speech Intelligibility," p. 629].

> Speech-enhancement algorithms tend to work better at high signal-to-noise ratios, but with high SNR the intelligibility is often very good to begin with. [. . .] [M]ost speech intelligibility and speech quality tests show that performance is better with steady, stationary noise like the background hum within a moving automobile or train, and performance is worse for fluctuating or time-varying noise such as speech babble or interfering music. Considerable effort continues in the intelligibility and enhancement fields, and both theoretical and practical breakthroughs are needed.

A less obvious factor in FSE is the nature of the speech that has been recorded. This is usually informal, unmonitored conversation, typically featuring overlapping utterances, moving speakers, whispering, or shouting—often in nonstandard varieties of English or other languages. Such speech can be difficult to understand even in a high-quality recording [7]. This sets limits on the clarity that could be achieved even if FSE were maximally effective.

Perhaps the most important factor, however, is precisely the uncertainty of the content. For nonforensic enhancing, the content is known or easily determinable. The enhancer's aim is to improve its quality (i.e., to make known content sound "better" in some relevant way). In FSE, by contrast, the aim is to assist a court in determining what the content is. This means improving not just the quality of the audio but also its intelligibility. That is a very different task and far harder for enhancing to achieve, especially to the standards required of the forensic sciences, which are responsible for ensuring that their methods are demonstrably and consistently capable of achieving the results claimed for them [8, 9]. As noted on the Audio Engineering Society's webpage, "Audio Forensics" [10]:

> As a practice, audio forensics is first and foremost a forensic science. This means that the factors important in all forensic disciplines are no less important here: standard practices, concepts of individualization, evidence handling and documentation, ethics, awareness of cognitive biases, clear and concise presentation of findings, and so on.

Finally, FSE is a high-stakes endeavor, capable of influencing the verdict in a criminal trial. Recognizing all these factors, well-qualified speech-enhancing experts often decline forensic casework or provide it with a low level of confidence. This leaves a huge demand for confident FSE, often met by poorly qualified practitioners using techniques they do not fully understand (online discussion forums such as "Sound on Sound" regularly feature requests for advice from practitioners struggling to fulfil the brief of an FSE assignment they have taken on without realizing how hard it is). Such work can be of very low standard.

This situation has brought an appropriate reaction from the responsible FSE community, which has sought to establish guidelines for best practice [11–13]. However, there is little incentive for practitioners to adhere to these guidelines, since their clients are often more concerned about criteria for admission in court than criteria for scientific rigor.

2 HOW IS ENHANCED FORENSIC AUDIO ADMITTED AND USED IN CRIMINAL TRIALS?

Despite the fact that FSE is considered by the law to be a forensic science, the criteria for admission can be very lax. Practitioners are required only to report the techniques they used, not to demonstrate that they have objectively improved the clarity of the audio. This is because the law assumes it is easy for lawyers, judges, and juries to determine whether enhanced audio is "clearer" than the original simply by listening to it. However, this assumption overlooks the role of a transcript.

Well-established research shows that the easiest way to make indistinct audio clearer, with no need to alter the audio itself, is to provide a transcript [7]. The reason is that the transcript "primes" listeners' perception, assisting them to hear the content. The danger, however, is that this priming effect can be just as powerful if the transcript is inaccurate, causing listeners to feel they clearly hear

words that are not really there [14]. Readers who doubt this are urged to reflect on experiences such as those offered in [15, 16].

In court, recorded speech evidence is always accompanied by a transcript. It might be assumed that the law insists that these transcripts are scientifically validated, but this is not so. Forensic transcription is not recognized by the law as a science, and transcripts are typically provided by investigators from the case—again, on the assumption that it is easy for lawyers, judges, and juries to evaluate a transcript's accuracy simply by checking it against the audio during the trial process. However, this checking is not always effective, as shown by multiple case studies [17].

Once listeners have a transcript, whether accurate or not, FSE has a minimal effect on the clarity of forensic audio. In particular, it is extremely unlikely to reverse the influence of a misleading transcript. Worse still, to the extent that FSE makes audio seem clearer, it can have the paradoxical effect of enhancing the credibility of an inaccurate transcript [18]. In this way, FSE has the potential to contribute to injustice, the exact opposite of what the FSE community intends.

3 A COMPLEX PROBLEM IN NEED OF AN EFFECTIVE SOLUTION

The first step to finding an effective solution for any complex problem is to clearly identify the originating cause of the problem. The new Scientific Working Group on Digital Evidence (SWGDE) "Best Practices" [13] offer valuable help in achieving this. First, they make a clear acknowledgment (SEC. 1.3) that enhancing techniques, even when used according to best practice, are not capable of making the true content of indistinct forensic audio clearer with the consistency, reliability, and accountability required by modern forensic science standards.

It is important to emphasize that the problem here is not the fact that FSE cannot achieve this (currently) impossible effect. Rather the problem is the courts' unrealistic expectation of what FSE can achieve—perhaps influenced by societal misconceptions [19, 20]. It is also important to emphasize that the fact that FSE cannot fulfill this unrealistic expectation does not detract from the important role that FSE can play in helping the courts to understand poor-quality forensic audio. It is, however, important for the FSE community to have a good understanding of how forensic speech recordings are used in court and to consider exactly how FSE can play a consistently useful role *within that context*.

Again, the SWGDE "Best Practices" provide a useful key. Their SEC. 1.3.3 acknowledges that objective intelligibility and quality metrics used by other branches of speech science [21] are unable to consistently and reliably determine whether FSE has made indistinct audio clearer on any given occasion (see [22] for an early account, [23] for a recent reminder in a broader context, and [24] for a case where alternative simplified metrics can even be preferred). In fact, like other expert treatments of FSE [11, 12], they emphasize the crucial role of practitioners *listening* to evaluate whether their enhancing has been successful (SEC. 4.4).

This brings back the topic of human speech perception. Of course, audio experts have technical listening skills that enable them to evaluate audio quality in fine-grained ways not accessible to naïve listeners. However, these skills are insufficient for evaluating "clarity" and "intelligibility," which are far more complex perceptual phenomena than usually recognized in audio and signal processing disciplines [25]. Over many decades, the linguistic and social sciences have brought attention to the essential role of the listener's language experience [26] and contextual expectations [27] in speech perception. These factors are even more important for indistinct forensic audio, whose crucial characteristic is its potential to sound clearly like something it is not.

In this context, defining intelligibility as "the proportion of a speaker's output that a listener can readily understand" (SEC. 5.1) instantly begs questions as to "Which listener?" and "Under what conditions?"—but these questions are not addressed at all by the SWGDE guidelines or by any other considerations of FSE that the authors are aware of. This is understandable, because these issues are generally considered outside the disciplinary expertise of audio specialists—which is precisely the reason this communication seeks to promote an interdisciplinary approach to evaluating FSE.

4 THE NEED FOR AN INNOVATIVE INTERDISCIPLINARY APPROACH

The law treats forensic speech enhancing and forensic transcription as two distinct ways to assist the courts in understanding indistinct speech recordings. However, they are really two sides of the same coin. Handling them separately reduces the value of each, to the point they can contribute to injustice, rather than to justice. Harnessing the power of both offers potential to deliver what the law really needs: reliable, accountable opinion evidence, to appropriate forensic science standards, capable of assisting the courts in determining the content of indistinct forensic audio or in determining that the audio content is not retrievable and recommending the evidence be excluded.

This communication aims to engage the interest of researchers from AES and beyond in an interdisciplinary project being pioneered by the Research Hub for Language in Forensic Evidence at the University of Melbourne, Australia. This project aims to achieve the needed theoretical and practical breakthroughs in the handling of forensic speech recordings, by combining expertise in both forensic speech enhancing and forensic transcription.

Soon the authors will launch an international survey to determine the current capability of FSE, and work to develop a high-level position statement setting realistic expectations regarding what FSE can achieve and, importantly, what simply cannot be achieved. From there, the authors hope to enlist collaboration from enhancing researchers and practitioners in developing end-to-end procedures for enhancing and transcribing poor-quality forensic speech recordings.

5 ABOUT THE AUTHORSHIP OF THIS COMMUNICATION

Following a webinar on this topic hosted by the International Speech Communication Association's Special Interest Group on Security and Privacy in Speech Communication [15], a self-selected subset of the audience formed an email discussion group. As well as exchanging valuable knowledge and ideas, members of this group conducted informal experiments on poor-quality forensic-like audio that were of great assistance in developing the arguments put forward in this communication.

6 REFERENCES

[1] D. Yu, Y. Gong, M. A. Picheny, et al., "Twenty-Five Years of Evolution in Speech and Language Processing," *IEEE Signal Proc. Mag.*, vol. 40, no. 5, pp. 27–39 (2023 Jul.). https://doi.org/10.1109/MSP.2023.3266155.

[2] D. O'Shaughnessy, "Speech Enhancement—A Review of Modern Methods," *IEEE Trans. Hum.-Mach. Syst.*, vol. 1, no. 1, pp. 110–120 (2024 Feb.). https://doi.org/10.1109/THMS.2023.3339663.

[3] M. A. Akeroyd, W. Bailey, J. Barker, et al., "The 2nd Clarity Enhancement Challenge for Hearing Aid Speech Intelligibility Enhancement: Overview and Outcomes," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5 (Rhodes Island, Greece) (2023 Jun.). https://doi.org/10.1109/ICASSP49357.2023.10094918.

[4] M. Cooke, S. King, M. Garnier, and V. Aubanel, "The Listening Talker: A Review of Human and Algorithmic Context-Induced Modifications of Speech," *Comput. Speech Lang.*, vol. 28, no. 2, pp. 543–571 (2014 Mar.). https://doi.org/10.1016/j.csl.2013.08.003.

[5] J. Richter, S. Welker, J.-M. Lemercier, et al., "Causal Diffusion Models for Generalized Speech Enhancement," *IEEE Open J. Signal Process.*, vol. 5, pp. 780–789 (2024 Mar.). https://doi.org/10.1109/OJSP.2024.3379070.

[6] AES, "Conference Report," *J. Audio Eng. Soc.*, vol. 62, no. 9, pp. 622–630 (2014 Jun.). https://www.aes.org/events/reports/54thConference.pdf.

[7] H. Fraser and D. Loakes, "Acoustic Injustice: The Experience of Listening to Indistinct Covert Recordings Presented as Evidence in Court," *Law Text Culture*, vol. 24, no. 1, pp. 405–429 (2020). https://ro.uow.edu.au/ltc/vol24/iss1/16.

[8] President's Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, (Executive Office of the President, Washington DC, USA, 2016). https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

[9] Law Commission of Great Britain, *Expert Evidence in Criminal Proceedings in England and Wales, Report No. 325,* (The Stationery Office, London, UK, 2011). https://lawcom.gov.uk/project/expert-evidence-in-criminal-proceedings/.

[10] AES, "Audio Forensics," https://aes2.org/audio-topics/audio-forensics-2/ (accessed Jul. 2, 2024).

[11] R. C. Maher, *Principles of Forensic Audio Analysis* (Springer, Cham, Switzerland, 2018). https://doi.org/10.1007/978-3-319-99453-6.

[12] J. Zjalic, *Digital Audio Forensics Fundamentals: From Capture to Courtroom* (Routledge, London, UK, 2021). https://doi.org/10.4324/9780429292200.

[13] Scientific Working Group on Digital Evidence (SWGDE), "Best Practices for the Enhancement of Digital Audio," Report 20-A-001-2.0 (2023 Jul.).

[14] H. Fraser, "'Assisting' Listeners to Hear Words That Aren't There: Dangers in Using Police Transcripts of Indistinct Covert Recordings," *Aust. J. Forensic Sci.*, vol. 50, no. 2, pp. 129–139 (2018 Apr.). https://doi.org/10.1080/00450618.2017.1340522.

[15] H. Fraser, "Enhancing Forensic Audio: What Works, What Doesn't, and How Can We Know?" presented at the *International Speech Communication Association Special Interest Group on Security and Privacy in Speech Communication (ISCA SIG-SPSC) Webinar* (Virtual) (2023 Jun.). https://blogs.unimelb.edu.au/language-forensics/2023/06/15/video-enhancing-forensic-audio/.

[16] H. Fraser, "Don't Believe Your Ears: 'Enhancing' Forensic Audio Can Mislead Juries in Criminal Trials," *The Conversation* (2019 Apr.). https://theconversation.com/dont-believe-your-ears-enhancing-forensic-audio-can-mislead-juries-in-criminal-trials-113844.

[17] H. Fraser and B. Stevenson, "The Power and Persistence of Contextual Priming: More Risks in Using Police Transcripts to Aid Jurors' Perception of Poor Quality Covert Recordings," *Int. J. Evid. Proof*, vol. 18, no. 3, pp. 205–229 (2014 Jul.). https://doi.org/10.1350/ijep.2014.18.3.453.

[18] H. Fraser, "Enhancing Forensic Audio: What Works, What Doesn't, and Why," *Griffith J. Law Hum. Dignity*, vol. 8, no. 1, pp. 85–102 (2020 Aug.).

[19] F. Rumsey, "Audio Forensics: Not an Episode From CSI," *J. Audio Eng. Soc.*, vol. 64, no. 6, pp. 440–444 (2016 Jun.).

[20] E. Ferragne, A. G. Talbot, M. Cecchini, et al., "Forensic Audio and Voice Analysis: TV Series Reinforce False Popular Beliefs," *Languages*, vol. 9, no. 2, paper 55, (2024 Feb.). https://doi.org/10.3390/languages9020055.

[21] Y. Feng and F. Chen, "Nonintrusive Objective Measurement of Speech Intelligibility: A Review of Methodology," *Biomed. Signal Process. Control.*, vol. 71, Part B, paper 103204 (2022 Jan.). https://doi.org/10.1016/j.bspc.2021.103204.

[22] D. Sharma, G. Hilkhuysen, N. D. Gaubitch, M. Brookes, and P. Naylor, "C-Qual—A Validation of PESQ Using Degradations Encountered in Forensic and Law Enforcement Audio," presented at the *Proceedings of the AES 39th International Conference: Audio Forensics: Practices and Challenges* (2010 Jun.), paper 8-1.

[23] D. de Oliveira, S. Welker, J. Richter, and T. Gerkmann, "The PESQetarian: On the Relevance of Goodhart's Law for Speech Enhancement," *arXiv preprint arXiv:2406.03460* (2024 Jun.) https://doi.org/10.48550/arXiv.2406.03460.

[24] A. Alexander, L. Gerlach, T. Coy, O. Forth, and F. Kelly, "Handling Real-World Challenges of Variable Speech Quality and Multiple Speakers in Forensic Automatic Speaker Recognition Using VOCALISE," in *Proceedings of the AES 8th International Conference on Audio Forensics* (2024 Jun.), paper 4.

[25] D. Chandler, "The Transmission Model of Communication," http://visual-memory.co.uk/daniel/Documents/short/trans.html (accessed Jul. 2, 2024).

[26] A. Cutler, *Native Listening: Language Experience and the Recognition of Spoken Words*, (MIT Press, Cambridge, MA, 2012). https://doi.org/10.7551/mitpress/9012.001.0001.

[27] D. J. Bruce, "The Effect of Listeners' Anticipations on the Intelligibility of Heard Speech," *Lang. Speech*, vol. 1, no. 2, pp. 79–97 (1958 Apr.). https://doi.org/10.1177/002383095800100202.

THE AUTHORS


Helen Fraser

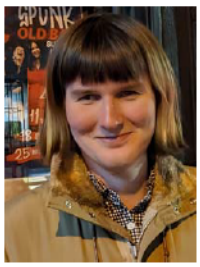

Vincent Aubanel


Rob Maher


Candy Olivia Mawalim


Xin Wang


Peter Počta


Emma Keith


Gérard Chollet


Karla Pizzi

Prof. Helen Fraser studied linguistics and phonetics at Macquarie University (Sydney) and The University of Edinburgh and then taught phonetics and related subjects for over twenty years at The University of New England (NSW). Following 10 years as an independent researcher and consultant, during which she undertook forensic case work and developed an influential research program in forensic transcription, Prof. Fraser is now Director of the Research Hub for Language in Forensic Evidence at The University of Melbourne.

•

Dr. Vincent Aubanel studied Computer Science, Music Technology and Phonetics before obtaining a Ph.D. in Phonetics from Aix-Marseille University (Laboratoire Parole et Langage, Aix-en-Provence, France) in 2011. Since then, he has held positions in Spain (Laslab, Universidad del País Vasco, Vitoria), Australia (MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Sydney), and France (GIPSA-lab, University of Grenoble-Alpes, Grenoble) before joining the Hub for Language in Forensic Evidence at the University of Melbourne in 2024 as a Research Fellow. His main interests are forensic speech, speech in noise, speech neuroscience, and speech rhythm.

•

Dr. Rob Maher is a Professor of Electrical and Computer Engineering (ECE), Montana State University in Bozeman, MT. He joined the ECE faculty in August 2002. He holds a B.S. degree from Washington University in St. Louis, M.S. degree from the University of Wisconsin-Madison, and Ph.D. from the University of Illinois-Urbana, all in Electrical Engineering. His research and teaching interests are in digital signal processing, with particular emphasis on applications in digital audio, audio forensic analysis, digital music synthesis, and acoustics. He is a Senior Member of IEEE, Associate Member of the American Academy of Forensic Sciences, and member of ASA, ASEE, Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi. Dr. Maher is a Fellow of the AES and currently serves as Associate Technical Editor and Deputy Editor-In-Chief of the AES journal.

•

Candy Olivia Mawalim received her B.S. in Computer Science from Institut Teknologi Bandung (ITB), Bandung, Indonesia. She received her M.S. and Ph.D. in the School of Information Science from the Japan Advanced Institute of Science and Technology (JAIST) in 2019 and 2022, respectively. She was selected as a Research Fellow for Young Scientists DC1 [Japan Society for

the Promotion of Science (JSPS)] in FY2020–2022. Since April 2022, she has worked as an assistant professor at the School of Information Science and Research Center for Biological Function and Sensory Information [Japan Advanced Institute of Science and Technology (JAIST)]. She is also on the education team of the International Speech Communication Association (ISCA) Special Interest Group of Security and Privacy in Speech Communication (SIG-SPSC) committee. Her main research interests are speech signal processing, hearing perception, voice privacy preservation, and machine learning.

•

Xin Wang received his Ph.D. degree from the SOKENDAI, Japan. He is a project associate professor at the National Institute of Informatics, Japan. He was among the organizing team of the Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenges and the VoicePrivacy initiatives. His research focuses on speech audio generation, text-to-speech synthesis, antispoofing, and other speech security and private related tasks.

•

Peter Počta received his M.Sc. and Ph.D. degrees in telecommunication from the Faculty of Electrical Engineering, University of Žilina, Slovakia, in 2004 and 2007, respectively. He is currently a Full Professor in the Department of Multimedia and Information-Communication Technology at theUniversity of Žilina. His research interests include multimedia quality assessment, multimedia and vehicular communications. He has published over 60 peer-reviewed papers in international journals and conferences including *IEEE Transactions on Broadcasting*, *Applied Acoustics*, *Speech Communication*, *Signal Processing: Image Communication, ACM Transactions onMultimedia Computing*, *Communications and Applications*, and the IEEE International Workshop on Multimedia Signal Processing (MMSP) and International Conference on Quality of Multimedia Experience (QoMEX). He also serves as an external reviewer for *IEEE Transactions on Multimedia*; *Digital Signal Processing*; *IEEE/ACM Transactions on Audio, Speech, and Language Processing*; and several conferences in the areas of multimedia quality, processing, and communication.

•

Emma Keith is a Ph.D. candidate in linguistics at the Australian National University. Her research interests include phonology, documentary linguistics, and forensic linguistics. Her Ph.D. research aims to document and describe Mentawai, an endangered language of Indonesia, including its phonology and morphosyntax.

•

Dr. Gérard Chollet was granted a Ph.D. in Computer Science and Linguistics from the University of California, Santa Barbara. He taught at Memphis State University and University of Florida before joining the Centre National de la Recherche Scientifique (CNRS). In 1981, he took in charge of the speech research group of Alcatel. In 1983, he joined a CNRS research unit at Ecole Nationale Supérieure des Télécommunications (ENST). In 1992, he participated to the development of the Idiap Research Institute, a research laboratory of the "Fondation Dalle Molle" in Martigny, Switzerland. From 1996 to 2012, he was back full time at ENST. CNRS decided in 2012 to grant him an emeritus status within the SAMOVAR laboratory (Télécom-SudParis). His main research interests are in phonetics, automatic audio-visual speech processing, spoken dialog systems, multimedia, pattern recognition, biometrics, privacy-preserving digital signal processing, speech pathology, and speech training aids.

•

Karla Pizzi is a Ph.D. candidate in cyber security at the Technical University in Munich. Her research interests include the security and privacy of speech-processing models, in particular automatic speech recognition and voice recognition. Besides this, Karla Pizzi works as a cyber security expert at Neodyme AG.