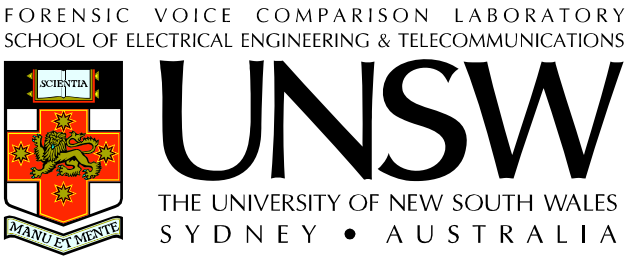

http://forensic.unsw.edu.au

Syllabus for ELEC9781
Forensic Voice Comparison and the Evaluation of Evidence

Document version of 29 July 2011

Course website: http://forensic.unsw.edu.au/ELEC9781.html

Times

First part: Monday 1 August – Friday 5 August 2011
Monday 8 August – Friday 12 August 2011
every day, 1:00pm–3:00pm

Second part: Thursday 25 August – Thursday 20 October 2011
every Thursday, 1:00pm–3:00pm

Location

Rm 319, Electrical Engineering Building (Kensington Campus map K17/G17)
University of New South Wales, UNSW Sydney, NSW 2052, Australia

Instructors

Geoffrey Stewart Morrison

geoff-morrison@forensic-voice-comparison.net

http://geoff-morrison.net http://forensic-voice-comparison.net

Dr Morrison is a Senior Research Fellow and the Director of the Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales (EE&T,

UNSW). He has a background in acoustic phonetics and statistical modelling. He has published a number of journal articles on forensic voice comparison and on the evaluation of forensic evidence. He is author of "Forensic Voice Comparison" in the *Expert Evidence* series (2010, Freckelton & Selby Eds). He is also an invited lecturer in the Judicial Phonetics Specialisation, Master of Phonetics and Phonology Programme, Spanish National Research Council / Menéndez Pelayo International University, an Adjunct Associate Professor, Department of Linguistics, University of Alberta, and Chair of the Forensic Acoustics Subcommittee, Acoustical Society of America. Dr Morrison is the lead researcher at UNSW for a project on *incorporation of forensic analysis techniques as part of an automatic speaker recognition system*. The project as a whole is lead by Dr Jason Pelecanos of IBM's T J Watson Research Center and funded by the Intelligence Advanced Research Projects Activity. Dr Morrison is also the lead researcher on an Australian Research Council Linkage Project on *making demonstrably valid and reliable forensic voice comparison a practical everyday reality in Australia*. The project includes researchers from EE&T and Law at UNSW, the Biometric Recognition Group at the Autonomous University of Madrid, and the Department of Forensic Science & Technology at the China Criminal Police University, and is a partnership with the Australian Federal Police, New South Wales Police, Queensland Police, National Institute of Forensic Science, Australasian Speech Science and Technology Association, and the Guardia Civil.

Julien Epps

j.epps@unsw.edu.au

http://www.ee.unsw.edu.au/staff/epps/profile.htm

Dr Epps is a Senior Lecturer, School of Electrical Engineering & Telecommunications, University of New South Wales. He is a member of the Forensic Voice Comparison Laboratory EE&T UNSW and a named researcher on the Australian Research Council Linkage Project on *making demonstrably valid and reliable forensic voice comparison a practical everyday reality in Australia*. He has co-authored a number of publications on speaker recognition and forensic voice comparison, and in 2008 led the UNSW team in the NIST Speaker Recognition Evaluation, as part of the highly successful I4U consortium submission (led by I^2R, Singapore). Dr Epps also leads other funded projects in emotion recognition and cognitive load classification from speech, and holds a joint appointment as a Senior Researcher with National ICT Australia.

Description

This course introduces the theory and practice of forensic voice comparison conducted within the new paradigm for the evaluation of forensic evidence.

The first part of the course (taught primarily by Morrison) focusses on the new paradigm for the evaluation of forensic evidence and its application to forensic voice comparison. The new paradigm is characterised by the quantitative implementation of the likelihood-ratio framework using databases and objective measurements, and empirical testing of the validity and reliability of forensic-comparison systems under conditions reflecting those of the case at trial. The likelihood-ratio framework is the logically correct framework for the evaluation of evidence and was adopted as standard for forensic DNA-profile comparison in the mid 1990s. Forensic voice comparison plays a leading rôle in the adoption of the new paradigm among other branches of forensic science.

The first part of the course should be of interest and accessible to a broad audience, with the concepts introduced applicable across all branches of forensic-comparison science and pertinent to the admissibility of forensic evidence in general. Interested lawyers, judges, police officers, and forensic scientists working in other branches of forensic science are welcome to audit this part of the course, which is taught in intensive mode.

The second part of the course (taught primarily by Epps) focusses on techniques from automatic speaker recognition and their application to forensic voice comparison. A number of automatic methods relevant to forensic voice comparison will be covered, including feature extraction for automatic speaker characterization (acoustic, dynamic, prosodic, linguistic), feature normalization, speaker modelling using Gaussian mixture models, speaker adaptation and GMM supervectors, channel compensation methods, and automatic segment selection.

The course will provide students with a solid grounding in the new paradigm for the evaluation of forensic evidence and advance students' technical skills and knowledge towards the state of the art in acoustic-phonetic and automatic approaches to forensic voice comparison.

Prerequisites

Auditors attending the first part of the course do not need to fulfill the prerequisites listed below.

Students taking the course for credit should normally fulfil the essential prerequisites (students lacking these prerequisites should consult with the instructors before enrolling):

ESSENTIAL:
– Knowledge of programming in MATLAB
– Knowledge of digital signal processing (e.g., ELEC3104 Digital Signal Processing)
– Knowledge of speech processing (ELEC9723 Speech Processing may be taken as a corequisite)

DESIRABLE:
– Knowledge of acoustic phonetics
– Knowledge of forensic science
– Knowledge of Bayesian statistics and statistical modelling
– Basic knowledge of linear algebra (e.g., matrix-vector operations, least squares)

Learning goals

The students will gain a solid understanding of the new paradigm for the evaluation of forensic evidence and its application to forensic voice comparison. They will be able to build a basic quantitative likelihood-ratio forensic-voice-comparison system and empirically test its validity and reliability.

Course structure – First part

The schedule for the first part of the course consists of two parallel tracks. One track is ordered thematically and evolves on the basis of discussions of readings. The other track focusses on the

cumulative development of knowledge and technical skills by means of a series of demonstrations and exercises in MATLAB. See the *course schedule - first part* document for the day to day schedule.

THEMATIC TRACK

– Paradigm shift

– Likelihood-ratio I: Philosophy

– Likelihood-ratio II: Calculation

– Evaluation I: Validity

– Evaluation II: Tippett plots

– Calibration and fusion

– Evaluation III: Reliability

– The alternative hypothesis

– Approaches to forensic voice comparison

– Resistance to the adoption of the new paradigm

TECHNICAL TRACK

– Logic and probability. Bayes' Theorem.

– Discrete probabilities. Continuous probabilities.

– Univariate normal distribution. Calculation of likelihood ratios.

– Multivariate normal distribution. The effect of correlation. Calculation of likelihood ratios.

– Gaussian mixture models. Calculation of likelihood ratios.

– Log-likelihood-ratio cost (C_{llr}).

– Logistic regression.

– Credible intervals.

– Training, optimisation, and evaluation databases. Cross validation.

– Formant-trajectory measurement.

– Parametric curves.

Course structure – Second part

The schedule for the second part of the course also consists of a reading-based thematic track and a more practical technical track. The former is based on readings, which will be discussed in the class, augmented by some presentation material. The latter is based on a combination of worked examples, tutorial-style exercises and lab work (using MATLAB).

THEMATIC TRACK

– Feature extraction for speaker characterization - acoustic and dynamic variants

Epps & Ambikairajah (2011); Kinnunen & Li (2010)

– Feature extraction for speaker characterization - prosodic and linguistic

Reynolds et al. (2003); Kockmann et al., (2010); Epps & Ambikairajah (2011); Kinnunen and Li (2010); Lu & Dang (2008); Thiruvaran et al. (2009)

– Feature normalization

Pelecanos & Sridharan (2001)

– Speaker modelling and Gaussian mixture models - what is being modelled and how GMMs should be used

Reynolds & Rose (1995); Bimbot et al. (2004); Kua et al., (2011); Kinnunen & Li (2010)

– Speaker adaptation and GMM supervectors

Gauvain & Lee (1994); Reynolds (1997); Bimbot et al. (2004); Campbell et al. (2005); Kinnunen & Li (2010)

– Channel compensation - NAP, JFA and i-vectors

Solomonoff et al. (2005); Campbell et al. (2006); Fauve et al. (2007); Kenny & Dumouchel (2004); Kenny et al. (2008); Dehak (2009)

– Segment selection for speaker characterization

Vair et al. (2007); Bocklet & Shriberg (2009); Huang et al. (2011); Pruthi & Espy-Wilson (2004)

TECHNICAL TRACK

– Implementation of selected features

– Gaussian mixture model parameter estimation using the Expectation Maximisation algorithm

– Modelling principles using Gaussian mixture models

– Implementation and visualization of universal background model and MAP adaptation

– Low-dimension implementation of nuisance attribute projection

– Implementation of segment selection algorithm

Teaching/learning approach

The teaching/learning approach for the course will consist primarily of discussions based on the reading and MATLAB exercises, but there will also be presentations by the instructors.

Each day we will cover part of the thematic track and/or part of the technical track. The instructor will

provide the students with a series of questions to guide them in reflecting on the most important aspects of the readings. The students should come to class ready to discuss these questions and other ideas which may emerge form the readings. “Ready to discuss” does not necessarily imply having all the answers at the beginning of the class, but we hope to have the answers at the end of the class or the end of the course.

Students are advised to read the essential readings for the first part of the course at least once before the beginning of the course, and then read the relevant ones again in preparation for each day of class. Students should be prepared to discuss the first day’s readings on the first day of class. Since the first part of the course is taught in intensive mode it is important to be well prepared and not to fall behind.

Assessment

Students are expected to complete the reading assignments and take an active part in class discussions.

Students are expected to attempt the Matlab exercises and self assess their progress by comparing their answers which the sample answers provided. The Matlab exercises will also be discusses in class but will not be graded.

There will be a short quiz on the last day of the first part of the course (Friday 12 August). This serves as a diagnostic of the learning of key concepts. It is intended to help the student assess their own learning and will not be graded.

Students will be asked to write a critical review of a published paper. The paper to be reviewed, and further details of the assignment will be announced on Friday 12 August (the last day of the first part of the course). The review should be from the perspective of the new paradigm for the evaluation of forensic evidence. Students may discuss the assignment with others, but should produce their own written review. A pdf of the review should be e-mailed to Morrison not later than 12:00 midday on Thursday 25 August (the first day of the second part of the course). This assignment will count for a third of the final grade.

Students will be asked to write a paper which demonstrates integration of knowledge and skills from both parts of the course. The paper should be in the format of an academic-conference paper or short refereed-journal paper. An outline of the paper including statement of the problem, literature review, and proposed methodology should be submitted as a pdf e-mailed to both Morrison and Epps not later than 12:00 midday on Thursday 6 October. The final version of the paper should be submitted as a pdf e-mailed to both Morrison and Epps not later than 12:00 midday on Thursday 10 November. This assignment will count for two thirds of the final grade.

Quality of writing, e.g., clarity of organisation and clarity of expression, will be taken into account in determining the grade assigned to each paper.

Registration

UNSW students taking the course for credit should register in ELEC9781 Special Topics in Electrical Engineering in Session 2 of 2011.

Persons wishing to audit the first part of the course should contact Morrison.

Credits

The course has 6 units of credit (equivalent to an expected workload is 10–12 hours per week throughout a 12 week session).

Academic honesty and plagiarism

Plagiarism is the unacknowledged use of others peoples work, including the copying of assignment works and laboratory results form other students. Plagiarism is considered a serious offence by the University and severe penalties may apply. For more information on plagiarism, please refer to:
http://www.lc.unsw.edu.au/plagiarism

Continual course improvement

Students are advised that the course in under constant revision in order to improve the learning outcomes of its students. Please forward any feedback on the course to the course instructors.

Administrative matters

On issues and procedures regarding such matters as special needs, equity and diversity, occupational health and safety, enrolment, rights and general expectation of students, please refer to the School policies, see:
http://scoff.ee.unsw.edu.au/

Graduate attributes

The course delivery methods and course content address a number of the UNSW graduate attributes, see:
http://learningandteaching.unsw.edu.au/content/userDocs/grad_attributes.pdf

Bibliography – First part

ESSENTIAL READINGS

Berger CEH, Buckleton J, Chanpod C, Evett IW, Jackson G (2011). Evidence evaluation: a response to the Court of Appeal judgment in R v T. *Science & Justice,* (published online 14 April 2011) doi:10.1016/j.scijus.2011.03.005.

Bolstad WM (2007). Logic, probability, and uncertainty. Ch. 4 in Bolstad WM, *Introduction to Bayesian Statistics* (2nd Ed.). Hoboken, NJ: Wiley.

Evett IW, *et al*. (2011). Expressing evaluative opinions: A position statement. *Science & Justice*, 51, 1–2. doi:10.1016/j.scijus.2011.01.002

French JP, Harrison P (2007). Position statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases. *International Journal of Speech, Language and the Law*, 14, 137–144. doi:10.1558/ijsll.v14i1.137

French P, Nolan F, Foulkes P, Harrison P, McDougall K (2010). The UK position statement on forensic speaker comparison: a rejoinder to Rose and Morrison. *International Journal of Speech, Language and the Law*, 17, 143–152. doi:10.1558/ijsll.v17i1.143

González-Rodríguez J, Rose P, Ramos D, Torre D, Ortega-García J (2007). Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, 15, 2104–2115. doi:10.1109/TASL.2007.902747

Morrison GS (2009a). Forensic voice comparison and the paradigm shift. *Science & Justice*, 49, 298–308. doi:10.1016/j.scijus.2009.09.002

Morrison GS (2009b). Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *Journal of the Acoustical Society of America*, 125, 2387–2397. doi:10.1121/1.3081384

Morrison GS (2010). Forensic voice comparison. Ch. 99 in Freckelton I, Selby H. (Eds.), *Expert Evidence*. Sydney, Australia: Thomson Reuters. http://expert-evidence.forensic-voice-comparison.net/

Morrison GS (2011a). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multvariate kernel density (MVKD) versus Gaussian mixture model – universal background model (GMM-UBM). *Speech Communication*, 53, 242–256. doi:10.1016/j.specom.2010.09.005

Morrison GS (2011b). Measuring the validity and reliability of forensic likelihood ratios. *Science & Justice*. (published online 14 April 2011) doi:10.1016/j.scijus.2011.03.002

Morrison GS (2012). The likelihood-ratio framework and forensic evidence in court: A response to R v T. *International Journal of Evidence & Proof*, 16(1) in press.

Morrison GS, Thiruvaran T, Epps J (2010). Estimating the precision of the likelihood-ratio output of a forensic-voice-comparison system. *Proceedings of Odyssey 2010: The Language and Speaker Recognition Workshop*, Brno, Czech Republic (pp. 63–70). http://www.isca-speech.org/archive_open/odyssey_2010/od10_012.html

R v T [2010] EWCA Crim 2439. http://www.bailii.org/ew/cases/EWCA/Crim/2010/2439.pdf

Rose P, Morrison GS (2009). A response to the UK position statement on forensic speaker comparison. *International Journal of Speech, Language and the Law*, 16, 139–163. doi:10.1558/ijsll.v16i1.139

Saks MJ, Koehler JJ (2005). The coming paradigm shift in forensic identification science. *Science*, 309, 892–895. doi:10.1126/science.1111565

RECOMMENDED READINGS

Aitken CGG, Lucy D (2004). Evaluation of trace evidence in the form of multivariate data. *Applied Statistics*, 54, 109–122. doi:10.1046/j.0035-9254.2003.05271.x

Aitken CGG, Taroni F (2004). *Statistics and the Evaluation of Forensic Evidence for Forensic Scientist* (2nd Ed.). Chichester, UK: Wiley.

Association of Forensic Science Providers (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science & Justice*, 49, 161–164. doi:10.1016/j.scijus.2009.07.004

Balding DJ (2005). *Weight-of-evidence for Forensic DNA Profiles*. Chichester, UK: Wiley.

Buckleton J (2005). A framework for interpreting evidence. Ch. 2 in Buckleton J, Triggs CM, Walsh SJ (Eds), *Forensic DNA Evidence Interpretation* (pp. 27–63). Boca Raton, FL: CRC.

Champod C, Meuwly D (2000). The inference of identity in forensic speaker recognition. *Speech Communication*, 31, 193–203. doi:10.1016/S0167-6393(99)00078-3

Evett IW (2009). Evaluation and professionalism. *Science & Justice*, 49, 159–160. doi:10.1016/j.scijus.2009.07.001

Evett IW (1991). Interpretation: A personal odyssey. In Aitken CGG, Stoney DA (Eds.), *The Use of Statistics in Forensic Science* (pp. 9–22). Chichester, UK: Ellis Horwood.

Hastie T, Tibshirani R, Friedman J (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd Ed.). New York: Springer.

http://www-stat.stanford.edu/~tibs/ElemStatLearn/

Kinnunen T, Li H (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, 52, 12–40. doi:10.1016/j.specom.2009.08.009

Reynolds DA, Quatieri TF, Dunn RB (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10, 19–24. doi:10.1126/science.1111565

Robertson B, Vignaux GA (1995). *Interpreting Evidence*. Chichester, UK: Wiley.

van Leeuwen DA, Brümmer N (2007). An introduction to application-independent evaluation of speaker recognition systems. In Müller C (Ed.), *Speaker Classification I: Selected Projects* (pp. 330–353). Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-540-74200-5_19

Bibliography – Second part

ESSENTIAL READINGS (DRAFT)

Bocklet, T. and Shriberg, E., "Speaker recognition using syllable-based constraints for cepstral frame selection", in *Proc. IEEE ICASSP*, 2009, pp. 4525–4528.

Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska, D., and Reynolds, D. A., "A tutorial ontext-independent speaker verification," *EURASIP Journal on Applied Signal Processing, Special issue on biometric signal processing*, 2004.

Campbell, W. M., Sturim, D. E., and Reynolds, D. A., "Supportvector machines using GMMsupervectors for speaker verification", *IEEE Signal Processing Letters*, vol. 13, no. 5, 2005, pp. 308–311.

Campbell, W., Sturim, D. E., Reynolds, D., and Solomonoff, A., "SVM based speaker verification usinga GMM supervector kernel and NAP variability compensation," in *Proc. IEEE ICASSP*, 2006, pp. 97–100.

Dehak, N., "Discriminative and Generative Approachesfor Long- and Short-Term Speaker Characteristics Modeling: Application to Speaker Verification", Ph.D. thesis, Ecole de Technologie Supérieure, Montréal, 2009.

Epps, J., and Ambikairajah, E., "Speech characterization and feature extraction for speaker recognition", in Li, G, and Li, H. (eds), *Advanced Topics in Biometrics*, World Scientific, 2011, to appear.

Fauve, B. G. B., Matrouf, D., Scheffer, N., Bonastre, J. F., and Mason, J. S. D., "State-of-the-art performance intext-independent speaker verification through open-sourcesoftware", *IEEE*

Transactions on Audio, Speech and Language Processing, vol. 15, no. 7, pp. 1960–1968, 2007.

Ferrer, L., Scheffer, N., and Shriberg, E., "A comparison of approaches for modeling prosodic features in speaker recognition", in *Proc. IEEE ICASSP*, pp. 4414–4417, Mar. 2010.

Gauvain, J., and Lee, C., "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 639–643, 1994.

Kinnunen, T., and Li, H., "An overview of text-independent speaker recognition: from features to supervectors," *Speech Communication*, vol. 52, 2010, pp. 12–40.

Kua, J. M. K., Epps, J., Nosratighods, M., Ambikairajah, E., and Choi, E. H. C., ?Using Clustering Comparison Measures for Speaker Recognition?, to appear in *Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Proc.* (Prague, Czech Republic), 2011.

Kenny, P., and Dumouchel, P., "Experiments in speaker verification using factor analysis likelihood ratios", in *Proc. Odyssey*, 2004, pp. 219–226.

Kenny, P.,Ouellet, P., Dehak, N., Gupta, V., and Dumouchel, P., "A study of inter-speaker variability in speaker verification", *IEEE Trans. Audio, Speech and Language Processing*, vol. 16, no. 5, pp. 980–988, July 2008. Available: http://www.crim.ca/perso/patrick.kenny

Kockmann, M., Burget, L., and Cernocky, J., "Investigations into prosodic syllable contour features for speaker recognition", in *Proc. IEEE ICASSP*, Mar. 2010, pp. 4418–4421.

Lu, X., and Dang, J., "An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification," *Speech Communication*, vol.50, 2008, pp. 312–22.

Pelecanos, J., and Sridharan, S., "Feature warping for robust speaker verification," in *Proc. Speaker Odyssey* (Crete, Greece), 2001, pp. I–640.

Pruthi, T., and Espy-Wilson, C. Y., "Acoustic parameters for automatic detection of nasal manner", *Speech Communication*, vol. 43, no. 3, August 2004, pp. 225–239.

Reynolds, D. A. and Rose, R. C.,"Robust text-independent speaker identification using Gaussian mixture models", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 3, no. 1, 1995, 72–83.

Reynolds, D. A., "Comparison of background normalization methods for text-independent speaker verification", in *Proc. European Conference on Speech Communication and Technology*, vol. 2, September 1997, pp. 963–966.

Reynolds, D., Andrews, W., Campbell, J., Navratil, J., Peskin, B., Adami, A.,Jin, Q., Klusacek, D., Abramson, J., Mihaescu, R., Godfrey, J., Jones, D., andXiang, B., "The SuperSID project:

exploiting high-level information for high accuracy speaker recognition", in *Proc. IEEE ICASSP*, 2003, pp. 4:784–787.

Solomonoff, A., Campbell, W. M., and Boardman, I., "Advances in channel compensation for SVM speaker recognition," in *Proc. IEEE ICASSP*, 2005, pp. 629–632.

Thiruvaran, T., Ambikairajah, E., and Epps, J., "Analysis of band structures for speaker-specific information in FM feature extraction", In *Proc. INTERSPEECH*, 2009, pp. 1111–1114.

Vair, C., Colibro, D., Castaldo, F., Dalmasso, E., and Laface, P., "Loquendo - Politecnico di Torino's 2006 NIST speaker recognition evaluation system," in *Proc. INTERSPEECH*, 2007, vol. 1, pp. 113–116.