

# Dr Korin Richmond

## Senior Research Fellow

School of Informatics, Informatics Forum, 10 Crichton Street  
Edinburgh, EH8 9AB, UK  
+44 (0)131 651 1769  
[korin@inf.ed.ac.uk](mailto:korin@inf.ed.ac.uk)  
<http://homepages.inf.ed.ac.uk/korin/>

### Statement Of Professional Interests And Research

The two factors that have motivated me most throughout my career are: i) fascination with human language, how it is produced and perceived; and ii) deep enthusiasm for machine learning, whereby we train computer models using representative data to perform human-like processing instead of programming them explicitly for a task. I have pursued these interests for over 20 years, most often combining the two for the development of speech-related technologies, but also in other applications of machine learning. My research can be categorised broadly into four areas:

Speech synthesis – I have worked extensively with state-of-the-art concatenative and statistical parametric methods for speech synthesis. A common theme has been to exploit an articulatory(-like) representation of speech. [outputs and impact: key publications [9, 13, 25, 50, 58]; Multisyn/Festival TTS engine software, used internationally; IEEE SPS Young Author Best Paper award 2010].

Articulatory modelling – The relationship between the articulatory and acoustic domains has interested me since my PhD. I have applied machine learning models for the acoustic-articulatory (inversion) mapping, the articulatory-acoustic (synthesis) mapping, and the acquisition and modelling of articulatory movement and shape data. [outputs and impact: key publications [1, 3, 13, 14, 25, 32, 38, 59]; release of [www.mngu0.org](http://www.mngu0.org) data set, used by >120 research groups worldwide; well-cited inversion mapping work - best reported method 2007-2015].

Pronunciation modelling – Particularly for speech synthesis and recognition: design and implementation of pronunciation lexica; machine learning for letter-to-sound conversion; regional accents; pronunciation variation; derivational morphology. [outputs and impact: key publications: [40, 51, 62]; Combilex lexicon licensed by leading international companies (e.g. Google, Samsung, Toshiba etc.) and universities (e.g. Cambridge, Sheffield etc.); consultancy for Illumina Digital Ltd (UK), for a BBC-funded project on synthesis of regional accents for a reading tuition tool].

Other machine learning applications – Besides speech, I have been keen to tackle problems in other areas where machine learning is of huge utility, e.g. tracking shapes in ultrasound video, and 2 Informatics MSc projects I proposed and supervised on i) validation of carbon fibre microscopy images; ii) magnetic field modelling. [outputs and impact: pub. [31]; both MSc projects awarded distinction].

### University Education

**1997-2002 — PH.D, CENTRE FOR SPEECH TECHNOLOGY RESEARCH (CSTR), UNIV. EDINBURGH**  
Thesis title: “*Estimating articulatory parameters from the acoustic speech signal*”; Advisors: Paul Taylor and Stephen Isard.

**1996-1997 — MASTER OF SCIENCE, UNIVERSITY OF EDINBURGH**  
Subject: Cognitive Science and Natural Language Processing; Dissertation topic: “*Compartmental modelling of stellate cells in the anteroventral cochlear nucleus, using realistic auditory nerve inputs*”.

**1991-1995 — MASTER OF ARTS (HONS), UNIVERSITY OF EDINBURGH**  
First class joint honours reading Linguistics and Russian.

## Academic Employment History

04/2001 — 03/2013	Full time	Research Fellow at the University of Edinburgh
03/2013 — 04/2016	Part time (80%)	Research Fellow at the University of Edinburgh
04/2016 —	Part time (80%)	Senior Research Fellow at the University of Edinburgh

## Funded research grants

### 2011-2014 RESEARCHER CO-INVESTIGATOR — “ULTRAX”

This multi-site EPSRC project (EP/I027696/1) concentrated on developing ultrasound as a tool for child speech therapy, by providing a simplified visual feedback display in realtime to guide the child's articulation attempts. Experience: ultrasound data acquisition and image processing algorithms, recursive Bayesian estimation, linear and nonlinear dimensionality reduction techniques. Leadership: originator and main author of technical aspects of initial grant proposal, and technical lead in project running.

### 2010-2011 RESEARCH FELLOW — “EMIME”

This multi-site EU-funded project (EC FP7 contract 213845) concerned mobile speech-to-speech translation, in particular aspects of cross-lingual personalization of speech synthesis and recognition components with rapid, automatic algorithms. Experience and responsibilities: a subproject evaluating multiple corpora of regional accents of English, and extending the Combilex multi-accent pronunciation lexicon to three new regional accents.

### 2010-2010 RESEARCH FELLOW — “BEETLE”

Funded by the US Office for Naval Research grant N00014-10-10085, this project was concerned with developing and evaluating an automated tutoring dialogue system. Experience and responsibilities: implemented an ICE (Internet Communications Engine - a framework for distributed computer systems) wrapper to allow the Festival speech synthesis system to be used as a component in the dialogue system; tailored a synthetic voice to suit the requirements for the dialogue system's spoken output in the project.

### 2009-2010 RESEARCH FELLOW (NAMED) — “EDINBURGH SPEECH PRODUCTION FACILITY”

A 3-year EPSRC project (EP/E01609x/1) to establish a facility to make articulatory recordings within the Informatics Forum at the University of Edinburgh. Two speakers can be recorded simultaneously with Carstens AG500 electromagnetic articulographs (EMA), allowing the recording and release of multiple 2-party dialogues. Experience and responsibilities: a 9-month self-proposed subproject to develop a faster and more robust method to process EMA data; successful approach applied recursive Bayesian estimation algorithms (e.g. Kalman Filtering, RTS smoothing). Leadership: originator and author of 9 month subproject of initial grant proposal, and technical lead in subproject running.

### 2006-2009 RESEARCHER CO-INVESTIGATOR — “PROBTTS”

EPSRC-funded project (grant EP/E027741/1) on using articulatory data in statistical parametric speech synthesis. Experience and responsibilities: development and extension of Hidden Markov model based speech synthesiser, data-driven articulatory-acoustic and acoustic-articulatory mappings, and supervision of Marie Curie visiting PhD fellows (incl. Zhenhua Ling, leading to a paper that was awarded the IEEE Best Young Author Paper Award, awarded 2011). Leadership: originator and main author of technical aspects of initial grant proposal, and technical lead in project running.

**2005-2006 RESEARCH FELLOW — “COMBILEX”**

An 18-month commercialisation project funded by the Scottish Government’s “Scottish Enterprise” agency to develop an advanced lexicon that encodes word pronunciations in all accents of English simultaneously. Experience: software development (predominantly Python, GUI programming, C++); database technology (PostgreSQL); finite state transducers, algorithm development and implementation (e.g. context dependent rewrite rules); lexicography; English derivational morphology and regional accent phonology. Leadership: technical lead in running of project during funding period, and sole responsibility for maintenance and further development of Combilex to date.

**2002-2005 RESEARCH FELLOW (NAMED) — “COUGAR”**

An EPSRC-funded project (GR/R94688/01) on exploiting articulatory speech measurements to improve the unit selection approach to speech synthesis. Experience and responsibilities: contributed as core developer of the *Festival* text-to-speech system and related *Edinburgh Speech Tools* software library (C++/Scheme programming); designed and implemented the *Multisyn* unit selection synthesis engine (rated joint third of 17 systems in the Blizzard Challenge international speech synthesis competition (2009)); speech synthesis technology generally; acoustic and articulatory data acquisition. Leadership: originator and co-author of technical aspects of initial grant proposal, and technical lead in project running.

**2001-2002 POSTGRADUATE RESEARCH ASSISTANT — “CONTINUOUS STATE DYNAMICAL SYSTEM MODELS FOR SPEECH RECOGNITION”**

Funded by EPSRC (GR/M75204/01), this project explored using linear dynamic system models for speech recognition, both using acoustic-only data, and also articulation as a hidden state space representation. Experience gained: automatic speech recognition technology, Kalman filtering, acoustic-articulatory relationship modelling.

**Funding procurement summary (as co-author/researcher co-investigator)**

Active	Funding	Title	Value
In preparation for submission to EPSRC (anticipated June 2016)		"Bridging the gap between language and speech"	£983,082 (FEC)
02/2011-07/2014	EPSRC (EP/I027696/1)	"Ultrax: Real-time Tongue Tracking for Speech Therapy"	£705,911 (FEC)
03/2011-09/2013	Royal Society of Edinburgh (RSE-NSFC International collaboration grant)	"Unified articulatory-acoustic modelling for flexible and controllable speech synthesis"	£11,639
12/2006-07/2010	EPSRC (EP/E01609X/1)	"An Edinburgh Speech Production Facility"	£619,576
11/2006-10/2009	EPSRC (EP/E027741/1)	"Data-driven Articulatory Modelling: Foundations for a New Generation of Speech Synthesis"	£286,857
07/2002-06/2005	EPSRC (GR/R94688/01)	"Concatenative Speech Synthesis Using Articulatory Information (Cougar)"	£181,097
Approximate total value (funded projects only):			£1.8M

**Research supervision**

- 3 PhD students:
  - João Cabral — graduated 2011, Ph.D awarded for thesis titled “HMM-based Speech Synthesis using an Acoustic Glottal Source Model”, now at University College Dublin, Ireland

- Qiong Hu, current student at CSTR - expected completion 2016, topic “Speech Synthesis with Sinusoidal Vocoding”
- External supervisor to Pertti Palo at Queen Margaret University - expected completion 2017, topic: ultrasound for speech reaction time monitoring and speech planning research
- 14 MSc student projects (solo- and co-supervised) on topics such as articulatory modelling, speech synthesis, pronunciation modelling, electromagnetic articulography coil position estimation, and machine learning for carbon fibre composite material validation.
- Projects of 6 visiting PhD students within the past 6 years: Chao Qin, Zhen-Hua Ling, Ingmar Steiner (all Marie-Curie Fellows), and Ming Lei, Blaise Potard, Maria Astrinaki. All projects have been of 3-8 months duration, and in areas relating to speech synthesis and recognition, as well as articulatory data acquisition and modelling.

## Teaching

Two years tutoring Linguistics 1&2 undergraduate courses at the University of Edinburgh.

Guest lecture on subject of Neural Networks for the Speech and Language Processing MSc course.

## Knowledge exchange and impact

- Developer and maintainer of an interactive public demonstration of the Festival text-to-speech synthesiser via the web. More than 1.5 million requests have been synthesised for visitors to the website to date. The demo is at <http://www.cstr.ed.ac.uk/projects/festival/onedemo.html>
- Responsible for the creation and release of the *mngu0* articulatory database - intended to remove a barrier to experimentation with articulatory data, and to facilitate direct comparison between different methods in the literature. Since initial announcement in 2011, this multimodal dataset (EMA, MRI, dental cast, video and audio data for a single male British English speaker) has been downloaded by more than 120 labs worldwide. More details available at project website: <http://www.mngu0.org>
- Combilex, an advanced speech technology lexicon developed with 18-months funding from the Scottish Enterprise “Proof-of-concept” scheme, has been commercially licensed by leading companies and universities worldwide (e.g. Google, Toshiba, Samsung, NICT (Japan), University of Cambridge, University of Sheffield etc.)
- 2 weeks consultancy for Illumina Digital Ltd (UK), contributing to a BBC-funded project requiring the synthesis of regional accents of English for a web-based reading tuition tool
- Multiple voices developed for the *Festival* text-to-speech system have been licensed commercially.
- Co-maintain CSTR’s web presence, with responsibility for CSTR’s publications database (and member listings) in particular

## Academic leadership, management and citizenship

Academic leadership and management experience:

- Co-organiser of Special Session on “Articulatory data acquisition and processing” at the annual Interspeech conference (the premier international conference in the speech field), France 2013
- Coordinator of CSTR’s weekly “Speak!” meeting, a forum to discuss audio/visual speech synthesis and to coordinate research effort (2004—2014)

**Membership of committees:**

- Member of organising committee for Ultrafest VI international workshop, held in November 2013 ([http://www.qmu.ac.uk/casl/conf/ultrafest\\_2013/default.htm](http://www.qmu.ac.uk/casl/conf/ultrafest_2013/default.htm))
- Serve on the international IEEE Speech and Language Technology Committee (elected 2011-2014, re-elected 2014-2017), designated as an expert in **speech synthesis**, also serving as the “Communications” subcommittee chair
- Management committee member for Marie Curie Research Training Site Programme “Edinburgh Speech Science and Technology (EdSST)”, run jointly between CSTR and Queen Margaret University (2006–2010)

**Consultancy:**

- 2 weeks consultancy for Illumina Digital Ltd (UK), contributing to a BBC-funded project requiring the synthesis of regional accents of English for a web-based reading tuition tool
- 1 week consultancy for iFlyTek Co. (China), processing of an English voice recording data set

**External recognition/esteem****Membership of societies:**

- Elevated to Senior Member of Institute of Electrical and Electronic Engineers (IEEE) Signal Processing Society (September 2014)
- Member, International Speech Communication Association (ISCA)

**Items of esteem:**

2010 IEEE Signal Processing Society’s Young Author Best Paper award, with Z.-H. Ling and J. Yamagishi

**Invited Talks:**

<b>2013</b>	<b>INVITED LECTURE AT THE INTERSPEECH CONFERENCE SATELLITE WORKSHOP “SPEECH PRODUCTION FOR AUTOMATIC SPEECH RECOGNITION (SPASR)”</b> “On Measuring and Estimating Speech Articulation”
<b>2013</b>	<b>TECHNICAL UNIVERSITY/FORSCHUNGSZENTRUM TELEKOMMUNIKATION, VIENNA, AUSTRIA</b> “Controllable speech synthesis”
<b>2012</b>	<b>GIPSA-LAB, GRENOBLE, FRANCE</b> "Articulatory speech technology"
<b>2011</b>	<b>UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA, HEFEI, CHINA</b> “Exploiting articulation in speech technology”
<b>2009</b>	<b>IPS, LUDWIG-MAXIMILIANS-UNIVERSITÄT, MUNICH, GERMANY</b> "Exploiting articulation for speech technology"
<b>2008</b>	<b>UK SPEECH MEETING, UNIVERSITY OF SURREY, GUILDFORD, UK</b> “Exploiting articulation in speech technology” (keynote)
<b>2007</b>	<b>LORIA, NANCY, FRANCE</b> "Estimating articulatory movements from the AG500 EMA and the acoustic speech waveform"
<b>2005</b>	<b>UNIVERSITY OF WEST BOHEMIA, PLZEN, CZECH REPUBLIC (TWO TALKS)</b> “Using articulatory information to improve unit select speech synthesis” & “An introduction to Festival and speech synthesis research at CSTR”
<b>2001</b>	<b>IMS, UNIVERSITÄT STUTTGART, GERMANY</b> “Accounting for non-uniqueness in the acoustic-to-articulatory inversion mapping”

Reviewing for all the primary journals and conferences in the field:

JOURNAL	CONFERENCE
<ul style="list-style-type: none"> <li>• Journal of the Acoustical Society of America</li> <li>• IEEE Transactions on Audio, Speech and Language Processing</li> <li>• Speech Communication</li> <li>• Computer Speech and Language</li> <li>• IEEE Signal Processing Letters</li> <li>• EURASIP Journal on Audio Speech and Music Processing</li> </ul>	<ul style="list-style-type: none"> <li>• Interspeech</li> <li>• IEEE ICASSP</li> <li>• SSW speech synthesis workshop</li> <li>• HLT-NAACL</li> <li>• NOLISP (Nonlinear speech processing)</li> <li>• Ultrafest</li> </ul>

2015 — invited review for Czech Science Foundation (national grant agency of the Czech Republic)

## Publications

	h-index	i10-index	total citations
Google Scholar statistics: (date 27.04.2016)	24	36	2285

KEY for publications listings below:

- i) \*\* — significant, as *included in Informatics' 2014 Research Excellence Framework entry*
- ii) \* — significant, for other reasons (incl. submission to previous RAE rounds)
- iii) K.Richmond — indicates role as main, or joint-main, originator

All publications are in peer-reviewed journals and conferences only.

All publications are available to download via: <http://www.cstr.ed.ac.uk/publications/users/korin.html>

## Books and theses published

\* [1] K. Richmond, "Estimating articulatory parameters from the acoustic speech signal," PhD thesis, The Centre for Speech Technology Research, Edinburgh University, 2002.

[2] K. Richmond, "A proposal for the compartmental modelling of stellate cells in the anteroventral cochlear nucleus, using realistic auditory nerve inputs," Master's thesis, Centre for Cognitive Science, University of Edinburgh, Sep. 1997.

## Articles published as sole author

\* [3] K. Richmond, "Preliminary inversion mapping results with a new EMA corpus," in Proc. Inter-speech, Brighton, UK, Sep. 2009, pp. 2835–2838.

[4] K. Richmond, "A multitask learning perspective on acoustic-articulatory inversion," in Proc. Inter-speech, Antwerp, Belgium, Aug. 2007.

[5] K. Richmond, "Trajectory mixture density networks with multiple mixtures for acoustic-articulatory inversion," in Advances in Nonlinear Speech Processing, International Conference on Non-Linear Speech Processing, NOLISP 2007, M. Chetouani, A. Hussain, B. Gas, M. Milgram, and J.-L. Zarader, Eds., ser. Lecture Notes in Computer Science, vol. 4885, Springer-Verlag Berlin Heidelberg, Dec. 2007, pp. 263–272. doi: 10.1007/978-3-540-77347-4\_23.

\* [6] K. Richmond, "A trajectory mixture density network for the acoustic-articulatory inversion mapping," in Proc. Interspeech, Pittsburgh, USA, Sep. 2006.

[7] K. Richmond, "Mixture density networks, human articulatory data and acoustic-to-articulatory inversion of continuous speech," in Proc. Workshop on Innovation in Speech Processing, Institute of Acoustics, Apr. 2001, pp. 259–276.

[8] K. Richmond, "Estimating velum height from acoustics during continuous speech," in Proc. Eurospeech, vol. 1, Budapest, Hungary, 1999, pp. 149–152.

#### Joint articles published

\* [9] K. Richmond and S. King, "Smooth talking: articulatory join costs for unit selection," 41st IEEE International Conference on Acoustics, Speech and Signal Processing, 2016.

\* [10] R. Dall, S. Brognaux, K. Richmond, C. Valentini-Botinhao, G. Eje Henter, J. Hirschberg, J. Yamagishi and S. King, "Testing the consistency assumption: pronunciation variant forced alignment in read and spontaneous speech synthesis," 41st IEEE International Conference on Acoustics, Speech and Signal Processing, 2016.

\* [11] Q. Hu, J. Yamagishi, K. Richmond, K. Subramanian, Y. Stylianou, "Initial investigation of speech synthesis based on complex-valued neural networks," 41st IEEE International Conference on Acoustics, Speech and Signal Processing, 2016.

[12] A. Hewer, S. Wuhler, I. Steiner, and K. Richmond, "Tongue mesh extraction from 3D MRI data of the human vocal tract," in Perspectives in Shape Analysis, M. Brey, A. M. Bruckstein, P. Maragos, and S. Wuhler, Eds., ser. Mathematics and Visualization. Springer, 2015.

\* [13] K. Richmond, Z.-H. Ling and J. Yamagishi, "The use of articulatory movement data in speech synthesis applications: an overview," Journal of Acoustical Science and Technology, vol. 36, no. 6, 2015, (12 page invited review).

\* [14] K. Richmond, J. Yamagishi and Z.-H. Ling, "Applications of articulatory movements based on machine learning," Journal of the Acoustical Society of Japan, vol. 70, no. 10, pp. 539–545, 2015.

[15] A. Hewer, I. Steiner, T. Bolkart, S. Wuhler, and K. Richmond, "A statistical shape space model of the palate surface trained on 3D MRI scans of the vocal tract," in Proceedings of the 18th International Congress of Phonetic Sciences, The Scottish Consortium for ICPhS 2015, Ed., Glasgow, United Kingdom, Aug. 2015, isbn: 978-0-85261-941-4.

\* [16] Q. Hu, Y. Stylianou, R. Maia, K. Richmond, and J. Yamagishi, "Methods for applying dynamic sinusoidal models to statistical parametric speech synthesis," in Proc. ICASSP, Brisbane, Australia, Apr. 2015.

\* [17] Q. Hu, Z. Wu, K. Richmond, J. Yamagishi, Y. Stylianou, and R. Maia, "Fusion of multiple parameterizations for DNN-based sinusoidal speech synthesis with multi-task learning," in Proc. Interspeech, Dresden, Germany, Sep. 2015.

\* [18] J. Cabral, K. Richmond, J. Yamagishi, and S. Renals, "Glottal spectral separation for speech synthesis," Selected Topics in Signal Processing, IEEE Journal of, vol. 8, no. 2, pp. 195–208, Apr. 2014. doi: 10.1109/JSTSP.2014.2307274.

[19] Q. Hu, Y. Stylianou, R. Maia, K. Richmond, J. Yamagishi, and J. Latorre, "An investigation of the application of dynamic sinusoidal models to statistical parametric speech synthesis," in Proc. Interspeech, Singapore, Sep. 2014, pp. 780–784.

\* [20] Q. Hu, Y. Stylianou, K. Richmond, R. Maia, J. Yamagishi, and J. Latorre, "A fixed dimension and perceptually based dynamic sinusoidal model of speech," in Proc. ICASSP, Florence, Italy, May 2014, pp. 6311–6315.

[21] M. Astrinaki, A. Moinet, J. Yamagishi, K. Richmond, Z.-H. Ling, S. King, and T. Dutoit, "Mage - reactive articulatory feature control of HMM-based parametric speech synthesis," in 8th ISCA Workshop on Speech Synthesis, Barcelona, Spain, Aug. 2013, pp. 227–231.

[22] M. Astrinaki, A. Moinet, J. Yamagishi, K. Richmond, Z.-H. Ling, S. King, and T. Dutoit, "Mage-HMM-based speech synthesis reactively controlled by the articulators," in 8th ISCA Workshop on Speech Synthesis, Barcelona, Spain, Aug. 2013, p. 243.

\* [23] C. Geng, A. Turk, J. M. Scobbie, C. Macmartin, P. Hoole, K. Richmond, A. Wrench, M. Pouplier, E. G. Bard, Z. Campbell, C. Dickie, E. Dubourg, W. Hardcastle, E. Kainada, S. King, R. Lickley, S. Nakai, S. Renals, K. White, and R. Wiegand, "Recording speech articulation in dialogue: evaluating a synchronized double electromagnetic articulography setup," *Journal of Phonetics*, vol. 41, no. 6, pp. 421–431, 2013. doi: 10.1016/j.wocn.2013.07.002.

[24] Q. Hu, K. Richmond, J. Yamagishi, and J. Latorre, "An experimental comparison of multiple vocoder types," in 8th ISCA Workshop on Speech Synthesis, Barcelona, Spain, Aug. 2013, pp. 155–160.

\*\* [25] Z. Ling, K. Richmond, and J. Yamagishi, "Articulatory control of HMM-based parametric speech synthesis using feature-space-switched multiple regression," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 207–219, Jan. 2013, issn: 1558-7916. doi: 10.1109/TASL.2012.2215600.

\* [26] K. Richmond, Z. Ling, J. Yamagishi, and B. Ura, "On the evaluation of inversion mapping performance in the acoustic domain," in *Proc. Interspeech*, Lyon, France, Aug. 2013.

\* [27] J. Scobbie, A. Turk, C. Geng, S. King, R. Lickley, and K. Richmond, "The Edinburgh speech production facility DoubleTalk corpus," in *Proc. Interspeech*, Lyon, France, Aug. 2013.

[28] I. Steiner, K. Richmond, and S. Ouni, "Speech animation using electromagnetic articulography as motion capture data," in *Proc. 12th International Conference on Auditory-Visual Speech Processing*, 2013, pp. 55–60.

[29] Z.-H. Ling, K. Richmond, and J. Yamagishi, "Vowel creation by articulatory control in HMM-based parametric speech synthesis," in *Proc. Interspeech*, Portland, Oregon, USA, Sep. 2012.

[30] Z. Ling, K. Richmond, and J. Yamagishi, "Vowel creation by articulatory control in HMM-based parametric speech synthesis," in *Proc. The Listening Talker Workshop*, Edinburgh, UK, May 2012, p. 72.

\*\* [31] K. Richmond and S. Renals, "Ultrax: an animated midsagittal vocal tract display for speech therapy," in *Proc. Interspeech*, Portland, Oregon, USA, Sep. 2012.

\*\* [32] I. Steiner, K. Richmond, I. Marshall, and C. D. Gray, "The magnetic resonance imaging subset of the mngu0 articulatory corpus," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, EL106–EL111, Jan. 2012. doi: 10.1121/1.3675459.

[33] I. Steiner, K. Richmond, and S. Ouni, "Using multimodal speech production data to evaluate articulatory animation for audiovisual speech synthesis," in *3rd International Symposium on Facial Analysis and Animation*, Vienna, Austria, 2012.

\* [34] B. Uria, I. Murray, S. Renals, and K. Richmond, "Deep architectures for articulatory inversion," in *Proc. Interspeech*, Portland, Oregon, USA, Sep. 2012.

[35] J. P. Cabral, S. Renals, J. Yamagishi, and K. Richmond, "HMM-based speech synthesiser using the LF-model of the glottal source," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 4704–4707. doi: 10.1109/ICASSP.2011.5947405.

\* [36] Z.-H. Ling, K. Richmond, and J. Yamagishi, "Feature-space transform tying in unified acoustic-articulatory modelling of articulatory control of HMM-based speech synthesis," in *Proc. Interspeech*, Florence, Italy, Aug. 2011, pp. 117–120.



[37] L. Ming, J. Yamagishi, K. Richmond, Z.-H. Ling, S. King, and L.-R. Dai, "Formant-controlled HMM-based speech synthesis," in Proc. Interspeech, Florence, Italy, Aug. 2011, pp. 2777–2780.

\* [38] K. Richmond, P. Hoole, and S. King, "Announcing the electromagnetic articulography (day 1) subset of the mngu0 articulatory corpus," in Proc. Interspeech, Florence, Italy, Aug. 2011, pp. 1505–1508.

[39] B. Uria, S. Renals, and K. Richmond, "A deep neural network for acoustic-articulatory speech inversion," in Proc. NIPS 2011 Workshop on Deep Learning and Unsupervised Feature Learning, Sierra Nevada, Spain, Dec. 2011.

\* [40] K. Richmond, R. Clark, and S. Fitt, "On generating Combilex pronunciations via morphological analysis," in Proc. Interspeech, Makuhari, Japan, Sep. 2010, pp. 1974–1977.

[41] A. Turk, J. Scobbie, C. Geng, B. Campbell, C. Dickie, E. Dubourg, E. G. Bard, W. Hardcastle, M. Hartinger, S. King, R. Lickley, C. Macmartin, S. Nakai, S. Renals, K. Richmond, S. Schaeffler, K. White, R. Wiegand, and A. Wrench, "An Edinburgh speech production facility," Poster presented at the 12th Conference on Laboratory Phonology, Albuquerque, New Mexico, Jul. 2010.

[42] J. Cabral, S. Renals, K. Richmond, and J. Yamagishi, "Transforming voice source parameters in a HMM-based speech synthesiser with glottal post-filtering," in Proc. 7th ISCA Speech Synthesis Workshop (SSW7), NICT/ATR, Kyoto, Japan, Sep. 2010, pp. 365–370.

[43] D. Felps, C. Geng, M. Berger, K. Richmond, and R. Gutierrez-Osuna, "Relying on critical articulators to estimate vocal tract spectra in an articulatory-acoustic database," in Proc. Interspeech, Sep. 2010, pp. 1990–1993.

[44] G. Hofer and K. Richmond, "Comparison of HMM and TMDN methods for lip synchronisation," in Proc. Interspeech, Makuhari, Japan, Sep. 2010, pp. 454–457.

[45] G. Hofer, K. Richmond, and M. Berger, "Lip synchronization by acoustic inversion," Poster at Siggraph 2010, Los Angeles, USA, 2010.

\* [46] Z.-H. Ling, K. Richmond, and J. Yamagishi, "An analysis of HMM-based prediction of articulatory movements," *Speech Communication*, vol. 52, no. 10, pp. 834–846, Oct. 2010, issn: 0167-6393. doi: 10.1016/j.specom.2010.06.006.

[47] Z.-H. Ling, K. Richmond, and J. Yamagishi, "HMM-based text-to-articulatory-movement prediction and analysis of critical articulators," in Proc. Interspeech, Makuhari, Japan, Sep. 2010, pp. 2194–2197.

[48] A. Turk, J. Scobbie, C. Geng, C. Macmartin, E. Bard, B. Campbell, C. Dickie, E. Dubourg, B. Hardcastle, P. Hoole, E. Kanaida, R. Lickley, S. Nakai, M. Pouplier, S. King, S. Renals, K. Richmond, S. Schaeffler, R. Wiegand, K. White, and A. Wrench, "The Edinburgh Speech Production Facility's articulatory corpus of spontaneous dialogue," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 2429–2429, 2010. doi: 10.1121/1.3508679.

[49] J. Cabral, S. Renals, K. Richmond, and J. Yamagishi, "HMM-based speech synthesis with an acoustic glottal source model," in Proc. The First Young Researchers Workshop in Speech Technology, Apr. 2009.

\*\* [50] Z. Ling, K. Richmond, J. Yamagishi, and R. Wang, "Integrating articulatory features into HMM-based parametric speech synthesis," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1171–1185, Aug. 2009, **IEEE SPS 2010 Young Author Best Paper Award**. doi: 10.1109/TASL.2009.2014796.

\* [51] K. Richmond, R. Clark, and S. Fitt, "Robust LTS rules with the Combilex speech technology lexicon," in Proc. Interspeech, Brighton, UK, Sep. 2009, pp. 1295–1298.

[52] I. Steiner and K. Richmond, "Towards unsupervised articulatory resynthesis of German utterances using EMA data," in Proc. Interspeech, Brighton, UK, Sep. 2009, pp. 2055–2058.

- [53] J. Cabral, S. Renals, K. Richmond, and J. Yamagishi, "Glottal spectral separation for parametric speech synthesis," in Proc. Interspeech, Brisbane, Australia, Sep. 2008, pp. 1829–1832.
- \* [54] Z.-H. Ling, K. Richmond, J. Yamagishi, and R.-H. Wang, "Articulatory control of HMM-based parametric speech synthesis driven by phonetic knowledge," in Proc. Interspeech, Brisbane, Australia, Sep. 2008, pp. 573–576.
- \* [55] C. Qin, M. Carreira-Perpinan, K. Richmond, A. Wrench, and S. Renals, "Predicting tongue shapes from a few landmark locations," in Proc. Interspeech, Brisbane, Australia, Sep. 2008, pp. 2306–2309.
- [56] I. Steiner and K. Richmond, "Generating gestural timing from EMA data using articulatory resynthesis," in Proc. 8th International Seminar on Speech Production, Strasbourg, France, Dec. 2008.
- [57] J. Cabral, S. Renals, K. Richmond, and J. Yamagishi, "Towards an improved modeling of the glottal source in statistical parametric speech synthesis," in Proc. of the 6th ISCA Workshop on Speech Synthesis, Bonn, Germany, 2007.
- \* [58] R. A. J. Clark, K. Richmond, and S. King, "Multisyn: open-domain unit selection for the Festival speech synthesis system," *Speech Communication*, vol. 49, no. 4, pp. 317–330, 2007. doi: 10.1016/j.specom.2007.01.014.
- \* [59] S. King, J. Frankel, K. Livescu, E. McDermott, K. Richmond, and M. Wester, "Speech production knowledge in automatic speech recognition," *Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 723–742, Feb. 2007.
- [60] K. Richmond, V. Strom, R. Clark, J. Yamagishi, and S. Fitt, "Festival multisyn voices for the 2007 blizzard challenge," in Proc. Blizzard Challenge Workshop (in Proc. SSW6), Bonn, Germany, Aug. 2007.
- [61] R. Clark, K. Richmond, V. Strom, and S. King, "Multisyn voices for the Blizzard Challenge 2006," in Proc. Blizzard Challenge Workshop (Interspeech Satellite), (<http://festvox.org/blizzard/blizzard2006.html>), Pittsburgh, USA, Sep. 2006.
- \* [62] S. Fitt and K. Richmond, "Redundancy and productivity in the speech technology lexicon - can we do better?," in Proc. Interspeech 2006, Sep. 2006.
- [63] R. A. J. Clark, K. Richmond, and S. King, "Multisyn voices from ARCTIC data for the Blizzard challenge," in Proc. Interspeech 2005, Sep. 2005.
- [64] G. Hofer, K. Richmond, and R. Clark, "Informed blending of databases for emotional speech synthesis," in Proc. Interspeech, Sep. 2005.
- \* [65] L. Onnis, P. Monaghan, K. Richmond, and N. Chater, "Phonology impacts segmentation in speech processing," *Journal of Memory and Language*, vol. 53, no. 2, pp. 225–237, 2005.
- \* [66] R. A. J. Clark, K. Richmond, and S. King, "Festival 2 – build your own general purpose unit selection speech synthesiser," in Proc. 5th ISCA workshop on speech synthesis, 2004.
- [67] D. Toney, D. Feinberg, and K. Richmond, "Acoustic features for profiling mobile users of conversational interfaces," in 6th International Symposium on Mobile Human-Computer Interaction - Mobile-HCI 2004, S. Brewster and M. Dunlop, Eds., Glasgow, Scotland: Springer, Sep. 2004, pp. 394–398.
- \* [68] K. Richmond, S. King, and P. Taylor, "Modelling the uncertainty in recovering articulation from acoustics," *Computer Speech and Language*, vol. 17, pp. 153–172, 2003.
- [69] J. Frankel, K. Richmond, S. King, and P. Taylor, "An automatic speech recognition system using neural networks and linear dynamic models to recover and model articulatory traces," in Proc. ICSLP, 2000.
- [70] S. King, P. Taylor, J. Frankel, and K. Richmond, "Speech recognition via phonetically-featured syllables," in PHONUS, vol. 5, Institute of Phonetics, University of the Saarland, 2000, pp. 15–34.

\* [71] A. Wrench and K. Richmond, "Continuous speech recognition using articulatory data," in Proc. ICSLP 2000, Beijing, China, 2000.

[72] K. Richmond, A. Smith, and E. Amitay, "Detecting subject boundaries within text: a language-independent statistical approach," in Proc. The Second Conference on Empirical Methods in Natural Language Processing, Brown University, Providence, USA, Aug. 1997, pp. 47–54.

#### **Other outputs/ non-text based material**

SOFTWARE: I have contributed as a core developer to CSTR's *Edinburgh Speech Tools* and *Festival* speech synthesis system since 2001. The web pages for these two software projects are:

[http://www.cstr.ed.ac.uk/research/projects/speech\\_tools](http://www.cstr.ed.ac.uk/research/projects/speech_tools)

<http://www.cstr.ed.ac.uk/research/projects/festival/>

DATA: I have recorded and collated a multimodal articulatory speech dataset. This unique dataset has been publicly released, with research-only licensing terms. The web site for this data resource is:

<http://www.mngu0.org>