

# Carlo Drioli

Dipartimento di Informatica  
Università degli Studi di Verona  
Strada Le grazie 15,  
37134 Verona, Italy

Phone (office): +39(045)8027968  
Fax: +39(045)  
Email: [carlo.drioli@univr.it](mailto:carlo.drioli@univr.it)  
Homepage: <http://profs.sci.univr.it/~drioli>

## Titoli di Studio

Dottorato in Ingegneria Elettronica e delle Telecomunicazioni, Università degli Studi di Padova, 14 Febbraio, 2003.

*Titolo della tesi:* Voice coding by means of physically-based models

*Supervisore:* Prof. Gian Antonio Mian,

Laurea in Ingegneria Elettronica, Università degli Studi di Padova, 8 Ottobre 1996, punteggio 107/110

*Titolo della tesi:* Applicazione degli Algoritmi Genetici e delle Reti Neurali alla sintesi per modelli fisici

*Relatori:* Prof. Giovanni De Poli, Ing. Davide Rocchesso

Diploma di Maturità Classica.

## Posizione Attuale

Assegnista di ricerca presso il Dipartimento di Informatica della Facoltà di Scienze Matematiche, Fisiche e Naturali, Università degli Studi di Verona.

Docente a contratto presso la Facoltà di Scienze Matematiche, Fisiche e Naturali, dell'Università degli Studi di Verona.

## Attività accademiche e di ricerca

Dall'Aprile 1997 all'Aprile 2000, e' stato assegnatario di una borsa di studio per attività di ricerca presso l'Università di Padova nell'ambito del progetto di ricerca "Cantieri Multimediali", promosso da Telecom Italia S.p.A. Nell'ambito del progetto, si è occupato di tecniche di analisi ed elaborazione del suono e della voce finalizzate al riconoscimento e alla modellazione del contenuto espressivo.

Dal 2000 al 2002 e' stato iscritto al Corso di Dottorato in Ingegneria Elettronica e delle Telecomunicazioni presso l'Università di Padova sotto la supervisione del Prof. Gian Antonio Mian. Ha conseguito il titolo di Dottore di Ricerca nel Febbraio 2003.

Dal Novembre 2000 al Gennaio 2003 ha partecipato, in qualità di collaboratore della ditta Generalmusic S.p.A, ad un progetto di ricerca dal titolo "Multisensory Expressive Gesture Application" (MEGA) finanziato dalla Commissione Europea nell'ambito del Quinto Programma Quadro. Il progetto mirava a sviluppare interfacce di interazione uomo-macchina basate sull'analisi dell'espressività del gesto. In questo progetto si è occupato di analisi e sintesi dei segnali audio e di analisi ed elaborazione della voce, e ha sviluppato moduli software per la piattaforma di interazione uomo-macchina "EyesWeb" del Dip. di Informatica e Sistemistica dell'Università di Genova.

Dal Settembre 2001 al Settembre 2002 ha svolto attività di ricerca nell'ambito del dottorato presso il Dipartimento di Speech, Music and Hearing (TMH) del Royal Institute of Technology di Stoccolma (KTH). La permanenza presso il KTH è stata parzialmente finanziata dalla Commissione Europea attraverso una borsa di studio "Marie Curie". Durante questo periodo ha approfondito lo studio di modelli fisici della sorgente vocale e diversi aspetti della modellazione, della sintesi e della trasmissione della voce.

Dal Febbraio 2003 al Settembre 2004 ha partecipato ad un progetto di ricerca dal titolo "Preparing Future Multisensorial Interaction Research" (PF-Star) finanziato dalla Commissione Europea nell'ambito del Quinto Programma Quadro. Il progetto mira a sviluppare interfacce di interazione uomo-macchina basate sull'analisi e la sintesi bimodale di parlato emotivo. In questo progetto si è occupato di analisi e sintesi dei segnali audio e di analisi e sintesi della voce.

Dal Luglio 2004 al Luglio 2005 ha avuto un rapporto formale di collaborazione con il Dip. di Informatica e Sistemistica dell'Università di Genova per lo sviluppo di moduli di elaborazione audio per la piattaforma di interazione uomo-macchina "EyesWeb". Le attività sono inserite nel contesto europeo del Sesto Programma Quadro con riferimento alla rete di eccellenza HUMAINE (Human-Machine Interaction Network on Emotion) ed al progetto IP Tai-Chi (Tangible Acoustic Interfaces for Computer-Human Interaction). Collabora tuttora con il DIST di Genova per la manutenzione e l'estensione della piattaforma EyesWeb.

Dal Febbraio 2003 al Settembre 2004 ha occupato una posizione con profilo di ricercatore, III livello professionale, con contratto a termine (ex. Art. 15) presso la Sezione di Padova di Fonetica e Dialettologia dell'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR.

E' stato responsabile del progetto CNR (Promozione Ricerca 2004) dal titolo "Indicizzazione audio per la lettura e l'ascolto della letteratura italiana", che ha riguardato lo studio e l'utilizzazione di sistemi di analisi della voce, sintesi vocale e di conversione Text-to-Speech per la fruizione di opere letterarie.

E' attualmente responsabile del progetto FSE (Fondo Sociale Europeo) dal titolo "Integrazione di funzioni vocali avanzate per l'interfaccia utente di applicazioni di centrale telefonica digitale", che riguarda lo studio e l'utilizzazione di sistemi di riconoscimento e sintesi del parlato in contesto di telefonia digitale.

Ha partecipato alle seguenti conferenze internazionali come relatore invitato: Forum Acusticum 2005, tenutosi a Budapest - Ungheria, nel Settembre 2005. Titolo dell'intervento: Physically oriented glottis models with inverse filtered waveform matching properties

Dall'Agosto 2005 è Assegnista di ricerca presso il Dipartimento di Informatica della Facoltà di Scienze Matematiche, Fisiche e Naturali, Università degli Studi di Verona.

Ha preso parte di recente ai seguenti progetti di ricerca: 1. Closing the Loop Of Sound Evaluation and Design (CLOSED) - progetto finanziato dalla Commissione Europea nell'ambito del sesto Programma Quadro. 2. Sound synthesis by physical models of the piano - progetto di ricerca congiunto tra il Dipartimento di Informatica della Università di Verona e Viscount International SpA (Joint Projects).

E' membro dell'IEEE, dell'ASA (Acoustic Society of America) e dell'ISCA (Int. Speech Communication Association)

## Attività didattiche

Dal 1997 al 2000 ha tenuto lezioni e seminari nell'ambito del corso "Sistemi di elaborazione per la musica" (Prof. G. De Poli), presso la Facoltà di Ingegneria dell'Università di Padova. Sono stati trattati argomenti inerenti a: elementi di psicoacustica, tecniche di sintesi dei suoni, modelli uditivi, tecniche di analisi e sintesi della voce. Ha collaborato alla stesura del capitolo "Elementi di acustica e psicoacustica" delle dispense del corso.

Nel Settembre 2000, ha tenuto un ciclo di lezioni (14 ore) per il Master universitario "Trasferimento delle tecnologie multimediali nel sistema della Piccola e Media Impresa del settore dei Beni Culturali" organizzato dalla Facoltà di Lettere e Filosofia dell'Università Federico II di Napoli. Nell'ambito del modulo "Suoni" ha trattato argomenti legati alla sintesi e alla elaborazione dei suoni e della voce.

Dal 2001 al 2003 ha partecipato alle lezioni del Corso di Perfezionamento in "Scienze Umane e Nuove Tecnologie" organizzato dal Dipartimento di Filosofia della Facoltà di Lettere e Filosofia dell'Università Federico II di Napoli. Gli interventi hanno come oggetto aspetti legati all'elaborazione audio e agli standard di codifica per i documenti multimediali.

Dall' a.a. 2000/2001 ha svolto attività di supervisione di tesi di laurea.

Per l'anno accademico 2005/2006 ha collaborato all'attività didattica per il corso "Elaborazione digitale di im-

magini e suoni - Suoni " e "Elaborazione digitale di immagini e suoni - Laboratorio Suoni " (Prof. D. Rocchesso) presso il Dipartimento di Informatica dell'Università di Verona.

Nel 2006 ha tenuto un ciclo di lezioni (18 ore) per il Master Universitario in Progettazione e gestione di sistemi di rete "Networked Multimedia Systems", presso il Dipartimento di Informatica dell'Università degli Studi di Verona. Gli argomenti trattati hanno riguardato la codifica e la compressione di audio e video, e le tecniche di correzione e mascheramento di errori per trasmissione su reti a pacchetto.

Nel Settembre 2008 ha tenuto un ciclo di lezioni riguardanti l'analisi del segnale vocale e il riconoscimento automatico del parlato nell'ambito della IV scuola estiva AISV dal titolo "Archivi di corpora vocali: conservazione, catalogazione, restauro audio e fruizione dei documenti sonori", svoltasi a Soriano nel Cimino, Viterbo.

Per l'anno accademico 2006/2007 ha ottenuto i seguenti incarichi di docenza a contratto presso la Facoltà di Scienze MM.FF.NN dell'Università degli Studi di Verona:

"Laboratorio di Informatica di Base", corso di Laurea in Informatica

"Laboratorio di Programmazione", corso di Laurea in Bioinformatica

Per l'anno accademico 2007/2008 ha ottenuto i seguenti incarichi di docenza a contratto presso la Facoltà di Scienze MM.FF.NN dell'Università degli Studi di Verona:

"Laboratorio di Metodi Informazionali", corso di Laurea in Bioinformatica

"Laboratorio di Programmazione", corso di Laurea in Bioinformatica

Per l'anno accademico 2008/2009 ha ottenuto i seguenti incarichi di docenza a contratto presso la Facoltà di Scienze MM.FF.NN dell'Università degli Studi di Verona:

"Laboratorio di Metodi Informazionali", corso di Laurea in Bioinformatica

"Laboratorio di Programmazione", corso di Laurea in Bioinformatica

Per l'anno accademico 2009/2010 ha ottenuto il seguente incarico di docenza a contratto presso la Facoltà di Scienze MM.FF.NN dell'Università degli Studi di Verona:

"Laboratorio di Programmazione", corso di Laurea in Bioinformatica

Per l'anno accademico 2010/2011 ha ottenuto i seguenti incarichi di docenza a contratto presso la Facoltà di Scienze MM.FF.NN dell'Università degli Studi di Verona:

"Elaborazione delle Immagini e dei Suoni - Modulo Suoni", corso di Laurea in Informatica

"Interazione Uomo-Macchina", corso di Laurea Magistrale in Ingegneria e Scienze Informatiche

## Attività professionali

Da Gennaio 2008 ha un rapporto di collaborazione formale con il Conservatorio di Trieste G. Tartini per lo sviluppo di un sistema di streaming audio video ad alta qualità e bassa latenza per applicazioni musicali interattive.

## Attività di servizio alla comunità scientifica

Ha fatto parte del Comitato Scientifico e Organizzativo del 1<sup>o</sup> Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce: "MISURA DEI PARAMETRI" aspetti tecnologici ed implicazioni nei modelli linguistici, tenutosi a Padova dal 2 al 4 Dicembre 2004.

Ha ricoperto il ruolo di "session chair" nell'ambito del Workshop internazionale MAVeBA 2003, tenutosi a Firenze nel Dicembre 2003.

Ha fatto parte del Technical Committee per l'organizzazione della conferenza DAFx-09, tenutosi a Como nel settembre 2009.

Ha collaborato in qualità di reviewer alla valutazione di articoli sottoposti alle seguenti riviste:

IEEE Transactions on Neural Networks  
IEEE Transactions on Audio, Speech and Language Processing  
Journal of the Acoustical Society of America  
Acustica/Acta Acustica  
Signal Processing  
Journal of Applied Signal Processing  
Pattern Recognition  
Computer Speech and Language  
Journal of Multimodal User Interfaces  
Journal of New Music Research

## Corsi specialistici frequentati

"Dynamics of Speech Production and Perception" organizzato da NATO International Scientific Exchange Programmes - Advanced Study Institute, June 24 to July 6, 2002, Il Ciocco, Italy German-French Summerschool on "Cognitive and physical models of speech production, perception and perception-production interaction" organizzato da ICP Grenoble , HUB & ZAS Berlin IPDS Kiel , ZAS Berlin , 19th-24th of September 2004, Lubmin, Germany.

## Lingue straniere

Ottima conoscenza dell'inglese, scritto e orale.  
Buona conoscenza del francese, scritto e orale.

## Interessi di ricerca

Gli interessi di ricerca attuali vertono sull'analisi e sulla elaborazione di voce e suoni, sui relativi aspetti orientati alla codifica di documenti multimediali, sulle problematiche relative all'interazione uomo-macchina basata su analisi e sintesi della voce, su tecniche di machine learning per l'elaborazione dei segnali e sul loro impiego nella definizione di nuovi modelli di sorgente per la codifica e la rappresentazione di fenomeni acustici. In particolare, nell'ambito del progetto "Cantieri Multimediali" si sono studiate tecniche di segmentazione musicale, segmentazione fonetica del canto, tecniche di analisi ed elaborazione spettrale. In questo contesto, si sono acquisite conoscenze relative all'analisi dei contenuti di documenti multimediali e ai principali standard di codifica per il multimedia. Nell'ambito della partecipazione ai progetti europei "MEGA" e "HUMAINE" si sono ulteriormente approfondite le problematiche relative all'analisi e al riconoscimento dei contenuti nei segnali audio nel contesto dell'interazione uomo-macchina e al design di software e di interazione. Nell'ambito del dottorato di ricerca sono stati approfonditi gli aspetti relativi alla sintesi ed alla codifica di suoni e voce mediante modelli fisici e kernel methods. Legati a queste ricerche sono gli interessi nel campo dell'identificazione di sistemi non lineari e delle tecniche di machine learning applicate alla elaborazione dei segnali. Nell'ambito della partecipazione al progetto europeo "Pf-Star" si sono affrontate problematiche relative alla analisi e alla sintesi della voce, con attenzione particolare alla sintesi del parlato emotivo. Nell'ambito della partecipazione al progetto europeo "CLOSED" si sono affrontate problematiche relative alla analisi e alla sintesi di fenomeni acustici derivanti da processi fluidodinamici. Ulteriori interessi sono, infine, nel campo delle teorie dei giochi e dei modelli di competizione e cooperazione. In questo ambito si sono condotte ricerche sull'uso di modelli di apprendimento di strategie in gruppi di giocatori.

## Pubblicazioni

### *Articoli su rivista e capitoli di libro*

- [1] M. Cristani, A. Pesarin, C. Drioli, A. Tavano, A. Perina, and V. Murino. Generative modeling and classification of dialogs by a low-level turn-taking feature. *Pattern Recognition*, 44(8):1785–1800, 2011.
- [2] P. Cosi and C. Drioli. *Emotions of the Human Voice*, K. Izdebski (Ed.), volume III: Culture and Perception, chapter LUCIA: a new emotive/expressive Italian talking head, pages 153–176. Plural Publishing, 2009.
- [3] A. Camurri, C. Drioli, B. Mazzarino, and G. Volpe. *Sound to Sense - Sense to Sound*, Pietro Polotti and Davide Rocchesso (Eds.), chapter Controlling Sound with Senses: Multimodal and Cross-Modal Approaches to Control in Interactive Systems, pages 243–278. Logos Verlag, Berlin, 2008.
- [4] C. Drioli and D. Rocchesso. On the use of Kernel-based methods in sound synthesis by physical modeling. *Numer. Algor.*, 45:315–329, 2007.
- [5] F. Avanzini, S. Maratea, and C. Drioli. Physiological control of low-dimensional glottal models with applications to voice source parameter matching. *Acta Acustica united with Acustica*, 92(Suppl.1):731–740, September 2006.
- [6] C. Drioli. A flow waveform-matched low-dimensional glottal model based on physical knowledge. *J. Acoust. Soc. Am.*, 117(5):3184–3195, May 2005.
- [7] S. Canazza, G. De Poli, C. Drioli, A. Rodà, and A. Vidolin. Modeling and control of expressiveness in music performance. *Proceedings of the IEEE*, 92(4):686–701, April 2004.
- [8] Magno Caldognetto E., Cosi P., Drioli C., Tisato G., and Cavicchio F. and. Visual and acoustic modifications of phonetic labial targets in emotive speech: Effects of the co-production of speech and emotions. *Speech Communication*, Vol. 44, October 2004:173–185, 2004.
- [9] C. Drioli and D. Rocchesso. Orthogonal least squares algorithm for the approximation of a map and its derivatives with a rbf network. *Signal Process.*, 83:283–296, February 2003.
- [10] C. Drioli and F. Avanzini. Hybrid parametric-physiological glottal modelling with application to voice quality assessment. *J. Medical Engineering and Physics*, 24(7-8):453–460, 2002.
- [11] C. Drioli. Radial basis function networks for conversion of sound spectra. *EURASIP J. Appl. Signal Process.*, 2001:36–44, January 2001.
- [12] S. Canazza, G. De Poli, C. Drioli, A. Rodà, and A. Vidolin. Audio morphing different expressive intentions for multimedia systems. *IEEE MultiMedia*, 7:79–83, July 2000.

### *Articoli in Atti di conferenze internazionali*

- [1] M. Cristani, A. Pesarin, C. Drioli, V. Murino, A. Rodà, M. Grapulin, and N. Sebe. Toward an automatically generated soundtrack from low-level cross-modal correlations for automotive scenarios. In *ACM Multimedia, Proceedings of the 18th International Conference on Multimedia*, pages 551–560, 2010.
- [2] C. Drioli and D. Rocchesso. Acoustic rendering of particle-based simulation of liquids in motion. In *Proceedings of the 12th International Conference on Digital Audio Effects - DAFx-09*, Como, Italy, September 2009.
- [3] C. Drioli, P. Polotti, D. Rocchesso, S. Delle Monache, K. Adiloglu, R. Annies, and K. Obermayer. Auditory representations as landmarks in the sound design space. In *Proceedings of the 6th Sound and Music Computing Conference (SMC09)*, Porto, Portugal, July 2009.
- [4] M. Cristani, A. Pesarin, C. Drioli, A. Tavano, A. Perina, and V. Murino. Auditory dialog analysis and understanding by generative modelling of interactional dynamics. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 103–109, June 2009.

- [5] K. Adiloglu, R. Annies, K. Obermayer, Y. Visell, and C. Drioli. Adaptive bottle. In *Proceedings of Int. Computer Music Conf. (ICMC2008)*, pages 24–29, Belfast, Ireland, August 2008.
- [6] C. Drioli and P. Cosi. Audio indexing for an interactive italian literature management system. In *INTER-SPEECH 2008, 9th Annual Conference of the International Speech Communication Association*, page 2170, 2008.
- [7] A. Pesarin, M. Cristani, V. Murino, C. Drioli, A. Perina, and A. Tavano. A statistical signature for automatic dialogue classification. In *19th International Conference on Pattern Recognition (ICPR2008)*, pages 1–4, 2008.
- [8] C. Drioli and F. Avanzini. Improved fold closure in mass-spring low dimensional glottal models. In *Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007)*, Florence, Italy, December 2007.
- [9] E. Marchetto, C. Drioli, and F. Avanzini. Inversion of a physical model of the vocal folds via dynamic programming techniques. In *Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007)*, Florence, Italy, December 2007.
- [10] G. Somnavilla, P. Cosi, C. Drioli, and G. Paci. Sms-festival - a new tts framework. In *Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007)*, Florence, Italy, December 2007.
- [11] M. Nicolao, C. Drioli, and P. Cosi. Voice GMM modelling for FESTIVAL/MBROLA emotive TTS synthesis. In *INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing*, Pittsburgh, PA, USA, September 2006.
- [12] F. Avanzini, S. Maratea, and C. Drioli. Physiological control of low-dimensional glottal models. In *Proc. of the 4th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, Florence, Italy, October 29 - 31 2005.
- [13] A. Camurri, P. Coletta, C. Drioli, A. Massari, and G. Volpe. Audio processing in a multimodal framework. In *paper n. 6390 in Proceedings of AES 118th Convention*, Barcelona, Spain, May 2005.
- [14] Drioli C., Tesser F., Tisato G., and Cosi P. Control of voice quality for emotional speech synthesis. In Cosi P., editor, *CD Rom Proceedings of AISV 2004, 1st Conference of Associazione Italiana di Scienze della Voce, Padova, Italy, December 2-4, 2004*, pages 789–798. EDK Editore s.r.l., Padova, 2005.
- [15] C. Drioli. Physically oriented glottis models with inverse filtered waveform matching properties. In *Proc. of the Forum Acusticum 2005 Conference*, pages 2749–2751, Budapest, Hungary, 29 Aug - 2 Sept 2005.
- [16] Magno Caldognetto E., Cavicchio F., Cosi P., Drioli G., and Tisato G. Parametri per lo studio delle modificazioni articolatorie del parlato emotivo. In P.Cosi, editor, *I Convegno AISV*, volume Atti del I Convegno della Societ  Italiana di Scienze della Voce, pages 441–470. Padova, EDK PRESS, 2005.
- [17] Tesser F., Cosi P., Drioli C., and Tisato G. and. Emotional festival-mbrola tts synthesis. In *CD Proceedings INTERSPEECH 2005 Lisbon, Portugal, 2005*, pages 505–508, 2005.
- [18] Tisato G., Cosi P., Drioli C., and Tesser F. Interface: a new tool for building emotive/expressive talking heads. In *CD Proceedings INTERSPEECH 2005 Lisbon, Portugal, 2005*, pages 781–784, 2005.
- [19] F. Tesser, P. Cosi, C. Drioli, and G. Tisato. Emotional festival-mbrola tts synthesis. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, pages 505–508, 2005.
- [20] G. Tisato, P. Cosi, C. Drioli, and F. Tesser. Interface: a new tool for building emotive/expressive talking heads. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, pages 781–784, 2005.
- [21] Tesser F., Cosi P., Drioli C., and Tisato G. Prosodic data-driven modelling of narrative style in festival tts. In *CDRom Proceedings of 5th ISCA Speech Synthesis Workshop, 14th-16th June 2004, Carnegie Mellon University, Pittsburgh USA, 2004*.

- [22] E. Magno Caldognetto, P. Cosi, C. Drioli, G. Tisato, and F. Cavicchio. Coproduction of speech and emotion: bimodal audio-visual changes of consonant and vowel labial targets. In *Proceedings of AVSP 03*, pages 209–214, S. Jorioz, France, September 4-7 2003.
- [23] C. Drioli. Synthesis of voiced sounds by means of waveform adaptive physical models. in *Proc. of Stockholm Music Acoustics Conference (SMAC)*, pages 377–380, 2003.
- [24] C. Drioli and F. Avanzini. Non-modal voice synthesis by low-dimensional physical models. in *Proc. 3rd Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, 2003.
- [25] C. Drioli, G. Tisato, P. Cosi, and F. Tesser. Emotions and voice quality: experiments with sinusoidal modeling. In *Proc. of Voice Quality: Functions Analysis and Synthesis (VOQUAL) Workshop*, pages 127–132, Geneva, 27-29 August 2003.
- [26] F. Boccardi and C. Drioli. Sound morphing with gaussian mixture models. In *Proc. 4th COST-G6 Conference on Digital Audio Effects (DAFX01)*, pages 44–48, Limerick, Ireland, December 10-13 2001.
- [27] S. Canazza, G. De Poli, C. Drioli, A. Roda, and A. Vidolin. Expressive morphing for interactive performance of musical scores. In *Proc. of the First IEEE International Conference on Web Delivering of Music (WEDELMUSIC)*, pages 116–122, Limerick, Ireland, November 23-24 2001.
- [28] F. Avanzini, C. Drioli, and P. Alku. Synthesis of the voice source using a physically-informed model of the glottis. In *Proc. of the Int. Symposium on musical acustics (ISMA)*, pages 31–34, Perugia, Italy, September 2001.
- [29] C. Drioli and F. Avanzini. A physically-informed model of the glottis with application to voice quality assessment. In *Proc. of the 2nd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, Florence, Italy, September 2001.
- [30] C. Drioli and F. Avanzini. Model-based synthesis and transformation of voiced sounds. In *Proc. 3rd COST G-6 Conf. on Digital Audio Effects (DAFX-00)*, pages 44–48, Verona, Italy, December 7-9 2000.
- [31] S. Zabarella and C. Drioli. Transformation of instrumental sound related noise by means of adaptive filtering techniques. In *Proc. 3rd COST G-6 Conf. on Digital Audio Effects (DAFX-00)*, pages 237–240, Verona, Italy, December 7-9 2000.
- [32] S. Canazza, G. De Poli, R. Di Federico, C. Drioli, and A. Roda. Symbolic and audio processing to change the expressive intention of a recorded music performance. In *Proc. 2nd COST-G6 Workshop on Digital Audio Effects (DAFX99)*, pages 1–4, Trondheim, Norway, December 1999.
- [33] C. Drioli. Radial basis function networks for conversion of sound spectra. In *Proc. 2nd COST-G6 Workshop on Digital Audio Effects (DAFX99)*, pages 9–12, Trondheim, Norway, December 1999.
- [34] R. Di Federico and C. Drioli. An integrated system for analysis-modification-resynthesis of singing. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, volume 2, pages 1254 –1259 vol.2, oct 1998.
- [35] C. Drioli and R. Di Federico. Toward an integrated sound analysis and processing framework for expressiveness rendering. In *Proc. International Computer Music Conference (ICMC)*, pages 175–178, Ann Arbor, Michigan, oct 1998.
- [36] C. Drioli and D. Rocchesso. Learning pseudo-physical models for sound synthesis and transformation. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, volume 2, pages 1085 –1090 vol.2, oct 1998.
- [37] C. Drioli and D. Rocchesso. A generalized musical-tone generator with application to sound compression and synthesis. In *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97) -Volume 1 - Volume 1*, pages 431–, Washington, DC, USA, 1997. IEEE Computer Society.

## Seminari

Nell'ambito del ciclo di seminari del Dipartimento di Informatica dell'Università degli studi di Verona, ha tenuto nell'aprile del 2005 un seminario dal titolo "Modeling of the glottal flow by physical models with waveform matching properties" e nel giugno del 2007 un seminario dal titolo "On the use of physical modeling and machine learning methods in sound and voice synthesis".

Nell'ambito del ciclo di seminari dell'Institut de la Communication Parlée (ICP/CNRS, Grenoble) ha tenuto nel marzo 2005 un seminario dal titolo "Modeling of the glottal flow by physical models with waveform matching properties".

Nell'ambito del ciclo di seminari KTH/TMH "Higher Music Acoustic seminar series" ha tenuto nell'Ottobre del 2001 un seminario dal titolo "Time-Frequency representation for sound modeling and transformation".

Nell'ambito dello stesso ciclo di seminari, ha tenuto nell'Aprile del 2002 un seminario dal titolo "Synthesis of the voice source by physically constrained glottis models".

Carlo Drioli, "Nonlinear Periodic prediction for sound transformation". Invited lecture presentata nell'ambito del Workshop "Le trasformazioni dei suoni", tenutosi a Firenze, presso il centro di ricerca Tempo Reale, 1997.

Carlo Drioli, Luca Stefanutti, "Simulation of N-Person Social Dilemma Games with Neural Networks" presentata al 28-mo European Mathematical Psychology Group Meeting (EMPG97), Nijmegen, Olanda, 1997.

Last updated: May 3, 2011

## Sommario dei temi e dei risultati della ricerca

### *Codifica e sintesi di audio e voce mediante modelli fisici*

La possibilità di rappresentare la voce e i suoni mediante modelli fisici della sorgente è oggetto di studio e di interesse per diversi motivi. Ha avuto, ed ha tuttora, un ruolo fondamentale nel processo di comprensione di fenomeni fisici complessi, come nel caso dell'acustica degli strumenti musicali e della produzione vocale, e costituisce un approccio potenzialmente vantaggioso ed efficiente per fini di sintesi, codifica e compressione dei segnali. L'attività di ricerca svolta in questo ambito ha riguardato dapprima l'acustica degli strumenti musicali, con lo studio di una classe di modelli generalizzati, ispirati agli schemi a guida d'onda (waveguide), per la rappresentazione di forme d'onda campionate da strumenti a corda e a fiato. Già da questa fase si è presa in considerazione la possibilità di unire la tecnica della modellazione fisica con tecniche di identificazione di serie temporali e di ricostruzione di dinamiche non lineari con lo scopo di realizzare modelli fisici dotati della capacità di "apprendere" alcune caratteristiche da esempi sonori reali. La ricerca è proseguita affrontando in un secondo tempo la codifica e sintesi della voce basata sulla modellazione fisica della sorgente di eccitazione glottale. Si è proseguita questa linea di ricerca approfondendo aspetti della modellazione e della identificazione parametrica, e sono stati presi in considerazione alcuni ambiti applicativi proponendo ad esempio schemi di analisi della voce basati sui modelli studiati. I modelli considerati per la rappresentazione della sorgente di eccitazione glottale, riconducibili alla classe di modelli a masse e molle di Ishizaka e Flanagan, sono caratterizzati da una struttura semplificata e dalla presenza di componenti nonlineari parametriche nella catena di feedback che consentono la riproduzione di forme d'onda arbitrarie. Le strutture proposte sono studiate in modo da mantenere le proprietà dell'oscillazione indotta da flusso tipica delle corde vocali. Questo schema è stato utilizzato per valutare schemi di analisi, trasformazione e sintesi della voce. I modelli proposti, pur conservando una struttura relativamente semplice ed economica dal punto di vista computazionale, presentano molte caratteristiche dinamiche del sistema glottale, come l'innescò dell'oscillazione all'aumentare della pressione polmonare o il variare della frequenza di oscillazione al variare della tensione delle corde vocali. Nelle versioni più semplici questi schemi sono basati su un modello di flusso ispirato alla legge di Bernoulli. Questa parte funzionale è stata in seguito estesa mediante componenti parametriche che permettono al modello di "apprendere" alcune caratteristiche timbriche da forme d'onda reali del flusso glottale, stimate attraverso tecniche di filtraggio inverso dell'onda di pressione alle labbra. Dopo una fase di identificazione, il modello è in grado di riprodurre un regime oscillatorio stabile caratterizzato dalla forma d'onda della serie temporale di training. Si è verificato come sia possibile controllare, in fase di risintesi, le caratteristiche della forma d'onda agendo sui parametri della parte fisica. La natura fisica di questi modelli permette di ottenere modificazioni realistiche anche della qualità della voce, riproducendo modi di fonazione non modali.

### *Audio spectral processing*

In numerosi algoritmi di elaborazione numerica dei segnali audio, quali il controllo dell'altezza del suono (pitch shifting), delle qualità vocali del parlatore (voice conversion), o dei parametri acustici (intensità, brillantezza, pitch) nella sintesi dei suoni, vi è la necessità di modelli versatili per l'elaborazione spettrale. In questo ambito è stato esplorato l'uso delle reti RBF e di modelli statistici a misture gaussiane (GMM) per modellare le trasformazioni spettrali implicate nel controllo di parametri fondamentali di suoni reali, come intensità e pitch, o nel controllo della qualità dell'emissione vocale nella sintesi del parlato emotivo. Le tecniche studiate sono state utilizzate in particolare nell'ambito dell'analisi e del controllo dell'espressività nelle performance musicali. Gli studi sull'espressività delle performance musicali hanno permesso di stabilire quali sono i parametri acustici e musicali attraverso cui è possibile controllare il contenuto espressivo, e di definire dei modelli matematici per il controllo dell'espressività mediante variazioni di tali parametri. In questo contesto, lo studio di un modello di elaborazione spettrale è servito a realizzare le elaborazioni dei suoni implicate nel controllo dell'espressività nelle performance musicali.

### *Tecniche di machine learning per l'elaborazione dei segnali*

Le reti RBF sono usate in numerose applicazioni per risolvere problemi di curve fitting e per l'identificazione di sistemi non lineari. Nelle mie ricerche, le reti RBF sono state usate per rappresentare nonlinearità generiche all'interno di schemi numerici ad ispirazione fisica. Ciò ha portato a dover affrontare problemi di stabilità e controllabilità tipici per sistemi nonlineari a catena chiusa. In questo ambito è stato proposto un algoritmo, basato sul metodo orthogonal least square (OLS), per il controllo puntuale della derivata. Sono stati approfonditi alcuni aspetti legati al

controllo della stabilità per una classe di sistemi dinamici definiti attraverso iterazione di funzioni nonlineari (mappe iterate). Per mappe iterate a un ingresso e una uscita si è studiata, in particolare, la possibilità di determinare il tipo di comportamento dinamico (stabilità, moto oscillatorio, moto caotico) attraverso la regolazione delle derivate della mappa in corrispondenza di punti notevoli quali i punti fissi.

### *HCI in real time systems*

Nell'ambito della partecipazione ai progetti europei "MEGA" e "HUMAINE" si sono approfondite le problematiche relative all'analisi e al riconoscimento dei contenuti nei segnali audio nel contesto dell'interazione uomo-macchina. Sono stati realizzati una serie di moduli di estensione per la piattaforma di interazione EyesWeb che integrano varie funzionalità di audio processing (front-end acustico, framework di analisi/sintesi sinusoidale) con le funzionalità di analisi del moto e riconoscimento del gesto presenti in EyesWeb.

### *Sintesi vocale emotiva*

La produzione del parlato in generale, e del parlato emotivo in particolare, è caratterizzata da un'ampia varietà di qualità dell'emissione vocale, intesa come il risultato delle varie modalità di fonazione possibili. Questo aspetto riveste un ruolo importante nella resa delle emozioni nella comunicazione verbale e sta assumendo importanza crescente nell'ambito dei sistemi di riconoscimento e sintesi del parlato. In questo contesto si inquadra la tematica di ricerca volta a definire algoritmi e sistemi di analisi e sintesi della voce finalizzati al controllo di parametri acustici di interesse per la simulazione di tipi di fonazione non modali (es., voce morbida, squillante, aspirata, sussurrata, rauca, tremante). Si è studiato come queste diverse qualità vengono utilizzate nella produzione del parlato emotivo e si sono realizzati dei moduli di elaborazione del segnale, integrati in sistemi per la sintesi vocale per concatenazione di difoni (FESTIVAL/Mbrola), che permettono il controllo della qualità vocale. Si sono affrontati infine gli aspetti legati alla definizione e alla realizzazione di linguaggi mark-up per il controllo delle caratteristiche della sintesi attraverso l'utilizzo di "tag" atti a specificare il tipo di fonazione o l'intenzione emotiva in contesti di animazione facciale (talking heads).