

Carlo Drioli

Department of Mathematics and Computer Science
University of Udine
Via delle Scienze 206,
33100 Udine, Italy

Phone (office): +39(0432)249814
Email: carlo.drioli@uniud.it
Homepage:
<http://www.dimi.uniud.it/Members/carlo.drioli>

Education

PhD in Electronic and Telecommunication Engineering, University of Padova, received February 14, 2003.

Dissertation Title: Voice coding by means of physically-based models

Supervisor: Prof. Gian Antonio Mian,

Laurea degree in Electronic Engineering, University of Padova, received 8 October 1996.

Dissertation: Applicazione degli Algoritmi Genetici e delle Reti Neurali alla sintesi per modelli fisici (Application of Genetic Algorithms and Neural Networks to synthesis by physical modeling)

Supervisors: Prof. Giovanni De Poli, Ing. Davide Rocchesso

Employment

Assistant Professor with the Department of Mathematics and Computer Science, University of Udine.

Contract Professor with the Faculty of Mathematical, Physical and Natural Sciences, University of Verona.

Research activities

From April 1997 to April 2000: assignee of a research scholarship at the University of Padova for the "Cantieri Multimediali" research project, sponsored by Telecom Italia SpA. Within this project, he was involved in analysis and processing of sound and voice, for the recognition and modeling of expressive content.

From 2000 to 2002: PhD student at the University of Padova, under the supervision of Prof. Gian Antonio Mian. He obtained his Ph.D. in February 2003.

From November 2000 to January 2003: contractor with the Generalmusic SpA company, within a research project titled "Multisensory Expressive Gesture Applications (MEGA), funded by the European Commission under the Fifth Framework Programme. The project aimed at developing human-machine interfaces based on the analysis of expressiveness of the gesture. In this project he was involved in analysis and synthesis of audio and voice processing, and developed software modules for the "EyesWeb" software, a human-machine interaction framework developed at the Dept. of Computer and Systems Engineering, University of Genoa.

From September 2001 to September 2002 he carried out doctoral research with the Department of Speech, Music and Hearing (TMH) of the Royal Institute of Technology in Stockholm (KTH). The stay at KTH was partially funded by the European Commission through a "Marie Curie" grant. During this period he deepened the study of physical models of the vocal source and various aspects of modeling, synthesis and voice transmission.

From February 2003 to September 2004 he held a Researcher position at the Phonetics and Dialectology division of the Institute of Cognitive Sciences and Technologies, Italian Research Council (ISTC-CNR).

From February 2003 to September 2004 he participated in a research project entitled "Preparing Future Multisensorial Interaction Research" (PF-Star) funded by the European Commission under the Fifth Framework Programme. The project aimed at developing human-machine interfaces based on the analysis and synthesis of bimodal emotional speech. Within this project he worked on analysis and synthesis of voice and speech.

In 2004, he was responsible for the a CNR project entitled "Audio Indexing for an Italian literature management system", involving the study and use of voice analysis and speech synthesis for the management of literary works within a Content Management System framework (CMS).

From July 2004 to July 2005 he had a formal collaboration with the Dept. of Computer Science and Systems Engineering, University of Genova, for the development of audio processing modules to be included in the "Eye-sWeb" framework. The activities were framed in the context of the European Sixth Framework Programme with reference to the network of excellence HUMAINE (Human-Machine Interaction Network on Emotion) and the IP project Tai-Chi (Tangible Acoustic Interfaces for Computer-Human Interaction).

From August 2005 to March 2011 he has been a Research Assistant at the Department of Computer Science of the University of Verona. During this period, he took part in various research projects, including:

1. Closing the Loop Of Sound Evaluation and Design (CLOSED) - funded by the European Commission under the Sixth Framework Programme.
2. Natural Interactive Walking (NIW) - funded by the European Commission under the Sixth Framework Programme.

The research related to these projects concerned various aspects of sound and voice processing and coding, signal processing based on physical description and kernel methods, sound and speech based interactive systems, multi-modal interaction.

From April 2010 to April 2011 he was responsible for an European Social Fund (ESF) project entitled "Integration of advanced voice features for interactive applications within the framework of digital telephony". The project concerned the study and use of speech recognition and synthesis in context of digital telephony and PBX systems.

From December 2011 he is an Assistant Professor with the Department of Mathematics and Computer Science, University of Udine.

Teaching

From 1997 to 2000 he held lectures and seminars within the course "Computer Systems for the Music " (Prof. G. De Poli), at the Faculty of Engineering, University of Padova. The topics covered include: elements of psychoacoustic, techniques for sound synthesis, auditory models, voice analysis and synthesis. He co-authored the chapter "Elementi di acustica e Psicoacustica" (Elements of acoustics and psychoacoustics) of the course handouts.

In September 2000, he held a series of lectures (14 hours) for the Master "Transfer of Multimedia Technologies to the Small and Medium Enterprises, in the field of Cultural Heritage", organized by the Department of Philosophy, University of Naples Federico II. Within the module "Sounds" he has dealt with topics related to audio coding standards, and to synthesis and processing of sound and voice.

From 2001 to 2003 he held part of the lessons of the Course "Human Sciences and New Technologies" organized by the Department of Philosophy, University Federico II of Naples. The lectures covered aspects such as audio processing and coding standard for multimedia documents.

In academic year 2005/2006 he has contributed occasional lectures for the course "Digital processing of images and sounds - Sounds " and "Digital processing of images and sounds - Sound Lab"(Prof. D. Rocchesso) at the Department of Computer Science, University of Verona.

In 2006 he held a series of lectures (18 hours) for the Master course "Networked Multimedia Systems", Department of Computer Science, University of Verona. Topics covered included the encoding and compression of audio and video, and techniques of masking and correction of errors for transmission over packet networks.

In September 2008 he held a series of lectures concerning the analysis of speech and automatic speech recognition, as part of the Fourth Summer School AISV entitled "Speech corpora: preservation, cataloguing, audio restoration and use of sound archives", held in Soriano nel Cimino, Viterbo.

In 2006/2007 he taught the following courses as a Contract Professor at the University of Verona:

"Laboratorio di Informatica di Base", (Foundations of Computer Science - Laboratory)

"Laboratorio di Programmazione", (Computer Programming - Laboratory)

In 2007/2008 he taught the following courses as a Contract Professor at the University of Verona:

"Laboratorio di Metodi Informazionali", (Foundations of Computer Science - Laboratory)

"Laboratorio di Programmazione", (Computer Programming - Laboratory)

In 2008/2009 he taught the following courses as a Contract Professor at the University of Verona:

"Laboratorio di Metodi Informazionali", (Foundations of Computer Science - Laboratory)

"Laboratorio di Programmazione", (Computer Programming - Laboratory)

In 2009/2010 he taught the following courses as a Contract Professor at the University of Verona:

"Laboratorio di Programmazione", (Computer Programming - Laboratory)

In 2010/2011 he taught the following courses as a Contract Professor at the University of Verona:

"Elaborazione delle Immagini e dei Suoni - Modulo Suoni", (Sound and Image Processing - Sounds)

"Interazione Uomo-Macchina", (Human Computer Interaction)

Presently (2011/2012) he is teaching the following courses as a Contract Professor at the University of Verona:

"Programmazione", (Computer Programming - Theory)

"Laboratorio di Programmazione", (Computer Programming - Laboratory)

and the following course at the University of Udine:

"Tecnologie Web", (Web Technologies)

Since 2000 he act as a supervisor for Laurea thesis dissertations.

Professional activities

Since January 2008 he has a formal collaboration with the Conservatory of Music G. Tartini, in Trieste, Italy, for the development of a system for high quality and low latency audio/video streaming ("LOLA"), oriented to interactive music performances and applications.

Programming skills

Very good MATLAB, and C/C++ programming skills. Good Java programming skills. He contributed to the development of the following software projects/packages:

1. LOLA (LOW LATency audio visual streaming system) project: a low latency, high quality audio/video transmission system for network musical performance and interaction.
<http://www.conservatorio.trieste.it/artistica/ricerca/progetto-lola-low-latency>
2. SDK (Sound Design Toolkit): ecologically-founded sound synthesis for Max and Pd.
<http://www.soundobject.org/SDT/>
3. EyesWeb project: a software framework for multimodal interactive performance and expressive gesture analysis.
<http://www.infomus.org/EywMain.html>

Memberships and affiliations

Member of the IEEE, ASA (Acoustic Society of America) and ISCA (Int. Speech Communication Association)

Research community service

He took part to the Scientific and Organizing Committee of the 1st National Conference of the AISV - Italian Association of Speech Sciences, held in Padova 2 to 4 December 2004.

He served as "Session Chair" in the International Workshop MAVEBA 2003, held in Florence in December 2003.

He took part to the Technical Committee for the organization of the conference DAFX-09, held in Como in September 2009.

He took part to the Program Committee for the organization of the conference Sound and Music Computing 2011 (SMC-11), to be held in Padova in July 2011.

He served as a reviewer for the following journals:

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

Specific courses attended

"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" organized by ICP Grenoble , HUB & ZAS Berlin IPDS Kiel , ZAS Berlin , 19th-24th of September 2004, Lubmin, Germany.

Foreign languages

Excellent knowledge of English, written and oral.

Good knowledge of French, written and oral.

Publications

Journal papers and book chapters

- [1] C. Drioli and D. Rocchesso. Acoustic rendering of particle-based simulation of liquids in motion. *Journal on Multimodal User Interfaces*, 5(3-4):187–195, 2012.
- [2] 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.
- [3] 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.
- [4] A. Camurri, C. Drioli, B. Mazarino, 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.
- [5] C. Drioli and D. Rocchesso. On the use of Kernel-based methods in sound synthesis by physical modeling. *Numer. Algor.*, 45:315–329, 2007.

- [6] 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.
- [7] C. Drioli. A flow waveform-matched low-dimensional glottal model based on physical knowledge. *J. Acoust. Soc. Am.*, 117(5):3184–3195, May 2005.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] C. Drioli. Radial basis function networks for conversion of sound spectra. *EURASIP J. Appl. Signal Process.*, 2001:36–44, January 2001.
- [13] 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.

Conference papers

- [1] C. Drioli and A. Calanca. Speech modeling and processing by low-dimensional dynamic glottal models. In *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, Portland, OR, USA, 2012.
- [2] C. Drioli and A. Calanca. Voice processing by dynamic glottal models with applications to speech enhancement. In *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*, Firenze, Italy, 2011.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] C. Drioli and P. Cosi. Audio indexing for an interactive italian literature management system. In *INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association*, page 2170, 2008.
- [9] 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.

- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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 Società Italiana di Scienze della Voce, pages 441–470. Padova, EDK PRESS, 2005.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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*.
- [24] 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. Jorjioz, France, September 4-7 2003.
- [25] 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.
- [26] 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.

- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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 acoustics (ISMA)*, pages 31–34, Perugia, Italy, September 2001.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.

Seminar presentations and invited speeches

Within the series of seminars of the Department of Computer Science, University of Verona, he held in April 2005 a seminar entitled "Modeling of the glottal flow by physical models with waveform matching properties" and in June 2007 a seminar entitled "On the use of physical modeling and machine learning methods in sound and voice synthesis."

Within the series of seminars of the Institut de la Communication Parlee (ICP / CNRS, Grenoble), he held in March 2005 a seminar entitled "Modeling of the glottal flow by physical models with waveform matching properties".

He attended the following international conferences as invited speaker: Forum Acusticum 2005, Budapest - Hungary, September 2005. Title: Physically oriented glottis models with inverse filtered waveform matching properties

During the same cycle of seminars, he held in April 2002 a seminar entitled "Synthesis of the voice source by glottis physically constrained models".

Within the series of KTH/TMH seminars "Higher Acoustic Music Seminar Series", he held in October 2001 a seminar entitled "Time-frequency representation for sound modeling and transformation."

During the Workshop "Le trasformazioni dei suoni" (sound transformations), in Florence at the research center "Tempo Reale", he held in 1997 an invited seminar titled "Nonlinear Periodic prediction for sound transformation".

Last updated: November 6, 2012

Summary or research themes and principal results

Coding and synthesis of speech and audio by physical models

The representation of voice and sounds using physical models of the source is an interesting research topic for several reasons. Physical modeling had and still has a major role in the process of understanding complex physical phenomena, such as the acoustics of musical instruments and vocal production, and is a potentially profitable and efficient approach for signal synthesis, coding and compression. The research undertaken in this field has focused first on the sound of musical instruments, with the study of a class of generalized models, inspired by the waveguide modeling approach, able to reproduce sampled waveforms from string and wind instruments. Since this early stage, the possibility of combining the physical modeling approach and the techniques for identification of time series and the reconstruction of nonlinear dynamics was taken into account, with the aim of creating physical models with the ability to "learn" features from real sound examples.

The research continued at a later date addressing coding and speech synthesis based on physical modeling of the glottal excitation source. This line of research has been pursued by investigating aspects of modeling and parameter identification, and by addressing various application areas such as voice analysis, voice quality assessment and control, voice and speech synthesis, speech enhancement. The models considered for the representation of the glottal excitation source, derived from the masses and springs models first proposed by Ishizaka and Flanagan, are characterized by a simplified structure and by the presence of parametric nonlinear components in the dynamic loop. The proposed structures are designed so as to preserve the typical flow-induced oscillation properties of the vocal cords, while providing data fitting properties.

Upon training, these models are able to reproduce stable oscillatory regimes characterized by the waveform of the training time series (glottal waveform estimates obtained by inverse filtering), and recent research activity on this topic concerned the possibility of performing signal transformations, i.e., time stretching, pitch shifting and voice quality modifications, by controlling the physically inspired parameters of the model.

Audio spectral processing

In many algorithms for digital audio signal processing, there is the need for versatile models oriented to spectral processing. This is the case, for example, of the control of the fundamental frequency (pitch shifting), the vocal characteristics or qualities of the speaker (voice conversion) or the control of acoustic parameters (e.g., intensity, brightness, pitch) in musical sound synthesis. In this field, the use of RBF networks and Gaussian Mixtures Models (GMM) has been explored, to model the spectral changes involved in the control of key parameters of real sounds, including intensity and pitch, or the control of voice quality in emotional speech synthesis. The techniques studied were used in particular in the analysis and control of expressiveness in musical performance. The studies related to the expressiveness of music performance have established which are the acoustic and musical parameters through which it is possible to control the expressive content, and to define mathematical models for the control of expressiveness through changes in these parameters.

Machine learning techniques for signal processing

Kernel Machines and RBF networks are used in many applications to solve problems of curve fitting and for the identification of nonlinear systems. As for this topic, the Kernel Machine method were used to represent nonlinearities in general physically inspired models of sounds and voice. This resulted in having to deal with problems of stability and controllability for dynamical nonlinear systems. With respect to this problem, an algorithm based on Orthogonal Least Square method (OLS), for precise control of the derivative, was proposed. This algorithm permitted to study some aspects of the stability control for a class of dynamical systems defined by iteration of nonlinear functions (iterated maps). For iterated maps with one input and one output in particular, the ability to determine the type of dynamic behavior (stability, oscillatory motion, chaotic motion) through the regulation of derivatives of the major points of the map at the fixed points, has been studied.

HCI in real time systems

During the participation to the European projects "Mega" and "Humaine", the problems related to the analysis and recognition of the audio signals content in the context of human computer interaction have been faced. This led to

a series of extension modules for the EyesWeb platform, integrating various features for audio processing (auditory models based acoustic front-end, sinusoidal modeling analysis / synthesis) with the motion analysis and gestures recognition capabilities of EyesWeb.

Emotional speech

The production of speech in general, and emotional speech in particular, is characterized by a wide variety of voice qualities, understood as the result of the various types of phonatory configuration. This aspect plays an important role in the rendering of emotions in verbal communication and is assuming increasing importance in the fields of automatic speech recognition and speech synthesis.

In this context, the research activity undertaken aimed at defining algorithms and methods for the control of acoustic parameters of interest for the simulation of non-modal phonation types (e.g., soft voice, loud, pressed, breathy, whispered, creaky). It was studied how these different qualities are used in the production of emotional speech, and signal processing algorithms for the control of the voice quality were designed and integrated in diphone speech synthesis tools (FESTIVAL / Mbrola). The issues related to development and implementation of mark-up language for controlling the characteristics of the synthesis through the use of "tagged" documents were also faced.

Also related to this research topic, some aspects of speech analysis were tackled which concerned the characterization and classification of dialogs, in the context of social signal processing.

Last updated: November 6, 2012