MUSAE LAB RESEARCH: FROM ANTHROPOMORPHIC SPEECH TECHNOLOGIES TO HUMAN-MACHINE INTERFACES AND HEALTH DIAGNOSTICS TOOLS

Tiago H. Falk
Institut national de la recherche scientifique, Centre Énergie, Matériaux, et Télécommunications
INRS-EMT, Université du Québec, Montréal, QC, Canada

Abstract

This paper introduces the Multimedia/Mutimodal Signal Analysis and Enhancement (MuSAE) Laboratory, located at the Institut national de la recherche scientifique (INRS-EMT), University of Quebec, in Montréal, Québec, Canada. The MuSAE Lab conducts award-winning interdisciplinary research at the crossroads of multimedia and biomedical signal processing, with the end-goal of developing innovative anthropomorphic technologies, intelligent human-machine interfaces, and next-generation health diagnostic tools. The ultimate aim of the paper is to highlight the expertise available at the MuSAE Lab, with the end goal of fostering new collaborations and partnerships with related professionals in Canada and abroad.

Keywords: anthropomorphic, speech processing, human-machine interface, health diagnostics, interdisciplinary

Résumé

Cet article présente le MuSAE Lab (Multimedia/Multimodal Signal Analysis and Enhancement – Laboratoire de traitement et de rehaussement de signaux multimédia et multimodaux), situé à l'Institut national de la recherche scientifique (INRS-ÉMT), Université du Québec, Montréal, Canada. La recherche interdisciplinaire effectuée au MuSAE Lab, maintes fois primée, combine le traitement de signaux multimédia et de signaux biomédicaux afin de développer des technologies anthropomorphiques novatrices, des interfaces humain-machine intelligentes et des outils diagnostiques de prochaine génération. Le but ultime de cet article est de mettre en relief l'expertise du MuSAE Lab, et d'ainsi développer de nouveaux partenariats et collaborations avec les professionnels et experts du domaine au Canada et à l'international.

Mots-clés: technologies anthropomorphiques, traitement de la parole, interface humain-machine, diagnostic médical

1 Introduction

The Multimedia Signal Analysis and Enhancement (MuSAE) Laboratory was founded in 2012 and is located at the Institut national de la recherche scientifique, Centre Énergie, Matériaux, Télécommunications (INRS-EMT), in Montréal, QC, Canada. At the MuSAE Lab, we conduct inter-disciplinary research at the crossroads of human-computer interaction, multimedia signal processing, and neuroscience. Unlike technologies in existence today, the human brain has the ability to repair itself; make on-the-spot decisions; learn and adapt; and integrate multiple modalities, all whilst consuming the same amount of energy as an electric light bulb. Notwithstanding, despite the rapid development of neuroimaging technologies that have allowed us to better understand the inner workings of the brain, limited advances have been observed in translating such knowledge into anthropomorphic technologies. ***We are conducting research to try to bridge this gap.***

2 Getting to Know the MuSAE Lab

2.1 Research Axes

We conduct research across three axes: anthropomorphic technologies, human-machine interfaces, and health diagnostics. For the purpose of this paper, focus will be placed on the acoustics-related projects within these axes.

B.1. Anthropomorphic multimedia technologies

Audio quality (and quality-of-experience, QoE) perception plays a significant role in the acceptability of new services and technologies. As such, service providers and developers often conduct subjective opinion tests. Subjective tests, however, are very expensive and time-consuming, thus objective quality metrics are needed. Existing metrics, however, overlook the cognitive processing stage involved in the human quality judgment process and, instead, use statistical data mining on large amounts of subjectively scored data. Within this axis, we have paired neuroimaging with subjective listening tests in order to gain insight into the human quality and quality-of-experience perception processes for synthesized and noisy speech. We have also integrated human influence factors (e.g., affective states, attention) into music and speech QoE models. The interested reader is referred to representative references [1-3]; a more complete list is available at [4].

B.2. Human-machine interaction (HMI)

It is known that far-field voice activated HMIs suffer from detrimental room acoustics artifacts (e.g., reverberation and noise) and from varying vocal efforts. Within this axis, several subprojects exist, including i) environment-robust

speaker identification, ii) vocal-effort robust speaker verification, iii) blind room acoustics characterization, iv) objective speech quality assessment, and v) enhancement of bone-conducted speech. The interested reader is referred to [5-7] for representative works and [4] for a complete list.

B.3. Health diagnostics

Within this axis, we have developed tools to i) diagnose autism risk in pre-verbal toddlers, ii) objectively characterize the intelligibility of dysarthric speakers, iii) tracheoesophageal speaker, and cochlear implant users, iv) characterize depression levels and v) enable hum-based assistive technologies. The interested reader is referred to [8-10] for representative works and [4] for a complete list.

2.2 Facilities and Available Equipment

Lab facilities feature a sound-proof data collection room with varying-size multimedia displays and an adjoining data analysis room with multi-core control and analysis PCs. Table 1 lists the equipment available in the MuSAE Lab.

Table 1: MuSAE Lab equipment list.

Category	Equipment	Manufacturer
Electroencephalography (EEG)	Active II (64-channel)	Biosemi
	Emotiv EEG	Emotiv
	Emotiv+	Emotiv
	MUSE	Interaxon
	Enobio (8-channel)	Neuroelectrics
Near-infrared spectroscopy (NIRS)	NIRScout (384 channels)	NIRx
	NIRSport (64 channels)	NIRx
Physiological signal monitors	Platinum sensor kit	Shimmer
	Q Sensor	Affectiva
	Hexoskin	Carré Technology
	Wireless BPM	iHealth
	Bioharness 3	Zephyr
Eye tracking	Eye Link 1000	SR Research
	ETG glasses	SMI
	EyeX	Tobii
Hearing assessment	AD629e audiometer	Interacoustics
Other	Kinect	Microsoft
	Rift Developer Kit	Oculus
	Ubi	UCIC
	Tesla K40 GPU	NVIDIA
	HD600 headphone	Sennheiser
	Saffire Pro 24 audio interface	Focusrite

2.3 Partnerships

Through partnerships with the McGill Centre for Interdisciplinary Research in Music Media and Technology, MuSAE Lab members have access to state-of-the-art recording facilities such as ITU-T standardized spatial audio labs, critical listening labs, and immersive presence labs. We also maintain a close relationship with several industry partners to assure that our developed tools can positively impact society, both technologically and economically.

2.4 Open Science

We are strong believers in open science and research. As such, the majority of our developed software tools are posted online on Github (https://github.com/MuSAELab/) as open-source for others to use and modify for research purposes. Our objective speech quality metric, for example, was used in the 2014 IEEE REVERB (REverberant Voice Enhancement and Recognition Benchmark) Challenge.

3 Conclusions

The ultimate goal of this paper was to introduce the MuSAE Lab and highlight the available expertise with the hope of leading to new partnerships and collaborations with like-minded professionals (both academic and industry), working in allied and/or complementary areas, not only in the Montréal and Québec regions, but also across Canada and abroad. The interested reader is invited to visit the MuSAE Lab's website (http://musaelab.ca) for more details.

Acknowledgment

The author wishes to acknowledge the support of NSERC, FQRNT, CFI, MDEIE, MELS, MEIE, Nuance Foundation, Google, NVIDIA, SYTACom, RQRV and DFATD.

References

[1] T. Falk and S. Möller, Towards Signal Based Instrumental Quality Diagnosis for Text-to-Speech Systems, IEEE Signal Processing Letters, Vol. 15, pp. 781-784, 2008.

[2] R. Gupta, H. Banville, I. Albuquerque, and T. Falk, Using fNIRS to Characterize Human Influential Factors: Towards Models of Quality of Experience Perception for Text-to-Speech Systems, in Proc. fNIRS Meeting, 2014.

[3] S. Arndt, J. Antons, R. Gupta, K. Laghari, R. Schleicher, S. Möller, and T. Falk, The Effects of Text-to-Speech System Quality on Emotional States and Frontal Alpha Band Power, in Proc. IEEE EMBS Neural Engineering Conference, pp. 489 – 492, 2013.

[4] http://musaelab.ca/publications

[5] T. Falk and W.-Y. Chan, Modulation Spectral Features for Robust Far-Field Speaker Identification, IEEE Trans. Audio Speech Lang. Process., Vol.18, No.1, pp. 90-100, Jan. 2010.

[6] M. Paja and T. Falk, Whispered Speech Detection in Noise Using Auditory-Inspired Modulation Spectrum Features, IEEE Signal Processing Letters, Vol. 20, No. 8, pp. 783-786, Aug. 2013.

[7] T. Falk and W.-Y. Chan, Temporal Dynamics for Blind Measurement of Room Acoustical Parameters, IEEE Trans. Instrum. Meas., Vol. 59, No. 4, pp. 978-989, April 2010.

[8] T. Falk, W.-Y. Chan, and F. Shein, Characterization of Vocal Source Excitation, Temporal Dynamics and Prosody for Dysarthric Speech Intelligibility Estimation, Speech Commun, Vol. 54, No. 3, pp. 622-631, June 2012.

[9] J. Santos and T. Falk, Updating the SRMR-CI Metric for Improved Intelligibility Prediction for Cochlear Implant Users, IEEE/ACM Trans. Audio, Speech, and Language Processing, Vol. 22, No. 12, pp. 2197-2206, Dec. 2014.

[10] J. Chan, T. Falk *et al*, Evaluation of a Vocal Cord Vibration Switch as an Alternative Access Pathway for an Individual with Hypotonic Cerebral Palsy – a Case Study, Disability and Rehabil.: Assistive Technol., Vol. 5, No. 1, pp. 69-78, Jan. 2010.