HAL
open science

Introduction to the Special Issue on Intrinsic Speech Variation

Renato de Mori, O. Deroo, Samuel Dupont, L. Fissore, P. Laface, A. Mertins, Christian Wellekens

To cite this version:

Renato de Mori, O. Deroo, Samuel Dupont, L. Fissore, P. Laface, et al.. Introduction to the Special Issue on Intrinsic Speech Variation. Speech Communication, 2007, 49 (10-11), pp.761. 10.1016/j.specom.2007.05.006 . hal-00499188

HAL Id: hal-00499188

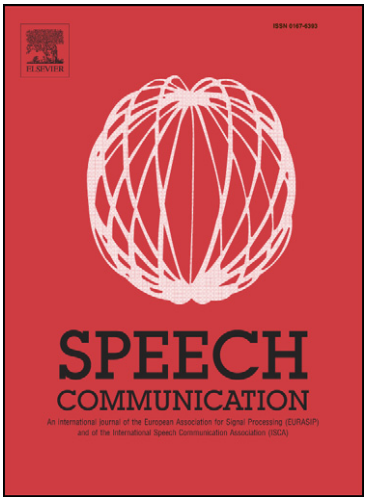
https://hal.science/hal-00499188v1

Submitted on 9 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript



Introduction to the Special Issue on Intrinsic Speech Variation

R. de Mori, O. Deroo, S. Dupont, L. Fissore, P. Laface, A. Mertins, Christian Wellekens

PII: S0167-6393(07)00098-2
DOI: 10.1016/j.specom.2007.05.006
Reference: SPECOM 1647

To appear in: *Speech Communication*

Received Date: 23 May 2007
Accepted Date: 23 May 2007

Please cite this article as: de Mori, R., Deroo, O., Dupont, S., Fissore, L., Laface, P., Mertins, A., Wellekens, C., Introduction to the Special Issue on Intrinsic Speech Variation, *Speech Communication* (2007), doi: 10.1016/j.specom.2007.05.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

1 Introduction

When asked about the progress of speech recognition, experts have to confess that in spite of available useful applications being carefully designed in order to avoid major problems, a long way is still to be walked to produce a system able to compare with human recognition and also another step is needed to reach understanding.

In the last years, huge efforts have been devoted to solve the problem of recognition in adverse environment and to desensitize recognizers versus the speech acquisition channel: both constitute extrinsic variabilities of speech. Dependence on speaker identity has also been extensively compensate by various techniques either in the signal space or in the model space. However a speaker is him/herself quite variable: stress, emotion, fatigue, health may lead to strong variations in speech effort, speech rate, accent that are important causes of recognition rate degradation. Submitting a proposal to European 6th Framework Program, the Guest editors expected to bring significant contributions to better understand the mechanisms of these intrinsic variations and to suggest and experiment solutions. The proposal was accepted under the name IST-2002-002034 project DIVINES.

The first task of the consortium was to write a state of the art which is reported in the first paper of this special issue. Later, the consortium organized a special session devoted to intrinsic variabilities during ICASSP 2006 in Toulouse (France) and also a workshop SRIV 2006 (Speech Recognition and Intrinsic Variabilities) as a satellite event of this conference. Both were quite successful and this special issue contains a selection of the contributions as well as papers from other researchers interested in this critical domain but not involved in DIVINES. Professor Richard Rose (Mc Gill University) took part to the project in the framework of an associated Canadian project.

In a second paper, Grimm et al. use fuzzy logic for the evaluation and estimation of emotions from speech in an emotional space which better describes the speaker state than simple classes like angry,sad,....

In a third paper, Casale et al. describe a genetic based method to select among the large number of existing acoustic feature sets which entries are best suited for detecting stressed/emotional states. They report experiments on the SUSAS (Speech Under Simulated and Actual Stress) database.

Articulatory features are advocated by many authors as a more flexible and parcimonious way to represent the acoustic signal in terms of phones and words as sequences of phones. The link between acoustic features and articulatory features is difficult to identify and it could become fuzzy under intrinsic variabililities. Scharenborg et al. analyze the behavior of two classifiers (MLP

ACCEPTED MANUSCRIPT

and SVM) for this difficult task.

Many adaptation methods have been proposed which modify either the acoustic vectors or the acoustic models. Clearly in HMM/AMM hybrid recognition systems, adaptation of the front-end MLP is quite unpractical but Gemello et al. describe a method where a layer of linear neurons can be efficiently adapted to compensate for intrinsic variations.

Accented speech seems to prevent recognition applications in international or even national public places. A brute force solution could be to develop an accent identifier and then to use a dedicated recognizer for each accent. A more clever solution is to create speech units trained on foreign data and to include phonological constraints. This is described with a paper by Bartkova and Jouvet.

Intra-speaker variability is a measure of the maturity of speech motor control and thus intrinsic variability is much greater in children's speech and decreases progressively when they grow up. Several specific adaptation methods are tested by Gerosa et al. on a large vocabulary task.

The project DIVINES took an end in January 2007. An important number of contributions were presented and published during its 3-year duration by the members of the consortium and this issue is only a very short selection of papers presented in Toulouse. The guest editors thank all authors including those who were not members of DIVINES, who submitted papers to this special issue. They also thank all anonymous reviewers.

R. De Mori, O.Deroo, S.Dupont, D. Jouvet,L.Fissore, P.Laface, A. Mertins, C.J. Wellekens
Guest editors

ACCEPTED MANUSCRIPT