Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 48 (2006) 1379–1398
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Band-pass filtering of the time sequences of spectral parameters for robust wireless speech recognition

J. Vicente-Peña *, A. Gallardo-Antolín, C. Peláez-Moreno, F. Díaz-de-María

Dpto. de Teoría de la Señal y Comunicaciones, EPS-Universidad Carlos III de Madrid, Avda. de la Universidad 30, 28911-Leganés, Madrid, Spain

Received 15 April 2005; received in revised form 25 July 2006; accepted 25 July 2006

Abstract

In this paper we address the problem of automatic speech recognition when wireless speech communication systems are involved. In this context, three main sources of distortion should be considered: acoustic environment, speech coding and transmission errors. Whilst the first one has already received a lot of attention, the last two deserve further investigation in our opinion. We have found out that band-pass filtering of the recognition features improves ASR performance when distortions due to these particular communication systems are present. Furthermore, we have evaluated two alternative configurations at different bit error rates (BER) typical of these channels: band-pass filtering the LP-MFCC parameters or a modification of the RASTA-PLP using a sharper low-pass section perform consistently better than LP-MFCC and RASTA-PLP, respectively.

Keywords: Robust speech recognition; Wireless speech recognition; Transmission errors; Modulation spectrum; RASTA-PLP

1. Introduction

Robustness in automatic speech recognition (ASR) systems has always been an extremely important issue since the first attempts to transfer this technology from research laboratories to real world applications. According to Junqua (2000) we can distinguish three sources of variability in speech that affect the performance of ASR systems: *task* and *speaker* is the first broad class, the second is the *acoustic environment* and last, *transducers* and *transmission channels*.

In this paper we are interested in dealing with the distortion produced by the new transmission channels that have emerged in voice transmission. This technology has experienced an enormous revolution in the past decade and still continues. These systems have evolved from the classical and sole transmission channel provided by public switched telephone network (PSTN) into a wide range of alternatives that include wireless cellular systems, VoIP, Bluetooth, wireless local and personal area networks of even a mixture of them.

* Corresponding author. Tel.: +34 91 624 9170; fax: +34 91 624 8749.
E-mail addresses: jvicente@tsc.uc3m.es (J. Vicente-Peña), gallardo@tsc.uc3m.es (A. Gallardo-Antolín), carmen@tsc.uc3m.es (C. Peláez-Moreno), fdiaz@tsc.uc3m.es (F. Díaz-de-María).

0167-6393/$ - see front matter
doi:10.1016/j.specom.2006.07.007

Besides, the pervasiveness of all these means of voice transmission has triggered the creation of multiple new information providing services that users can access through these networks. These services can greatly benefit from the use of automatic dialog systems for which an improved performance of the ASR subsystem over the particular underlying transmission channels can significantly reduce the need to resort to a human operator in many situations.

In this context, our work focuses on improving the robustness of ASR systems that are accessed through a wireless network. Thus, in this scenario the speech signal is transmitted through the corresponding wireless standard channel and is recognized at a remote server. This is not the only approach to this problem. Either embedded or distributed speech recognition (DSR) face up the same problem from a different point of view. Though not considered in this paper these alternatives are briefly reviewed and their drawbacks and advantages compared with those of the option considered here.

In this paper, we pay attention to two typical sources of distortion of wireless channels: lossy speech coding and transmission errors. Our work is inspired in previous works that suggested the filtering of the modulation spectrum of the speech features to deal with channel-distorted or noisy speech (Hermansky and Morgan, 1994; Hanson and Applebaum, 1993; Nadeu et al., 1997 are good examples). We have applied and adapted these ideas to the distortions typical of wireless speech communications.

As a starting point, we consider two well-known parameter sets, namely: MFCC and LP-MFCC. Further on, we focus on LP-MFCC since our experiments reveal that it performs better than MFCC in presence of coding distortion and transmission errors. We also compare our proposal with RASTA-PLP (Hermansky and Morgan, 1994), a well-known filtering-based parameter set.

We show, conceptually and experimentally, that a band-pass filtering of the time sequences of the spectral parameters is beneficial to deal with distortions due to transmission errors. Specifically, we suggest two configurations: the first one, called BPF-LP-MFCC, consists on a band-pass filtering of the LP-MFCC parameters; the second one is a modified version of RASTA-PLP, called M-RASTA-PLP, using a sharper low-pass section. In both cases, we obtain significant improvements with respect to LP-MFCC or the original RASTA, respectively, when transmission errors are considered.

The paper is organized as follows: Section 2 presents the problem of ASR in wireless communication systems; Section 3 describes the previous works on filtering the spectral parameters and discusses the reasons (either given by other authors in other contexts, or presented in this paper for wireless speech) for which we propose to improve and adapt this technique to the wireless speech communication scenario; Section 4 describes the experimental setup, the baseline systems, and the experimental assessment of the filtering-based proposed techniques in comparison to well-known robust parameterization methods; finally, conclusions and directions of further work are summarized in Section 5.

2. ASR in wireless environments

The enormous success of the wireless cellular systems makes the analysis of the distortion caused by them a relevant issue of research. With this purpose, we can identify the main sources of distortion originated by these systems that affect the performance of speech recognizers as

- *Acoustic environment*: Though strictly speaking this is not a distortion caused by the wireless system itself we have included this category into the classification to reinforce the idea that the wireless nature of these networks have broaden dramatically the variety of situations or acoustic environments in which voice is likely to be originated. Therefore, though indirectly, it poses a new challenge on the speech recognition systems.
- *Speech coding distortion*: The wireless bandwidth is a very expensive resource due to the increasing number of emergent wireless services that has only made worse the saturation that already existed in the radio-electric spectrum. Therefore, to optimize the productivity of the spectral bands that allow the transmission using electronic devices of mass production, extremely smart bandwidth sharing protocols have been devised. As part of these efforts to maximize the utilization of the spectrum the use of medium and low-rate speech coders plays a fundamental role in the feasibility of these networks in the market place. This aggressive compression of the speech signal produces a distortion that damages the speech recognizer operation.