Chapter 19
Neuromorphic Speech Processing: Objectives and Methods

Pedro Gómez-Vilda
Universidad Politécnica de Madrid, Spain

José Manuel Ferrández-Vicente
Universidad Politécnica de Madrid, Spain

Victoria Rodellar-Biarge
Universidad Politécnica de Madrid, Spain

Roberto Fernández-Baillo
Universidad Politécnica de Madrid, Spain

Agustín Álvarez-Marquina
Universidad Politécnica de Madrid, Spain

Rafael Martínez-Olalla
Universidad Politécnica de Madrid, Spain

Víctor Nieto-Lluis
Universidad Politécnica de Madrid, Spain

Luis Miguel Mazaira-Fernández
Universidad Politécnica de Madrid, Spain

Cristina Muñoz-Mulas
Universidad Politécnica de Madrid, Spain

ABSTRACT

Current trends in the search for improvements in well-established technologies imitating human abilities, as speech perception, try to find inspiration in the explanation of certain capabilities hidden in the natural system which are not yet well understood. A typical case is that of speech recognition, where the semantic gap going from spectral time-frequency representations to the symbolic translation into phonemes and words, and the construction of morpho-syntactic and semantic structures find many hidden phenomena not well understood yet. The present chapter is intended to explore some of these facts at a simplifying level under two points of view: that of top-down analysis provided from speech perception, and the symmetric from bottom-up synthesis provided by the biological architecture of auditory pathways. An application-driven design of a Neuromorphic Speech Processing Architecture is presented and its performance analyzed. Simulation details provided by a parallel implementation of the architecture in a supercomputer will be also shown and discussed.

DOI: 10.4018/978-1-61520-919-4.ch019
INTRODUCTION

Cognitive Speech Processing is a new concept in which traditional speech processing tasks such as speech recognition, speaker identification, language identification or emotion detection are examined under the point of view of speech processing in the auditory pathways and higher centers to contribute new ideas or paradigms directly inspired in the natural processing, in a first step to better understand how the real system processes speech, and in a second step to improve current processing results. In doing so, a preliminary review of speech processing in the auditory centers is mandatory. This is done under two different points of view: to the light of current knowledge in neurophysiology of the auditory pathways (Physiological Level: PhL) and under the perceptual point of view (Psychoacoustical or Perceptual Level: PsL). This approach is funded in the facts that neurophysiological knowledge is far from being complete, many important questions being still open, and responses are to be complemented from knowledge in the PsL. As the PsL can only reveal observations from the systemic behavioral response, the methodological approach must rely in the formulation of work hypotheses to be contrasted against physiological evidence by means of computer simulations. The general structure of the chapter will respond to this methodological reflection. First most relevant knowledge for the study from well-established facts at the PsL will be reviewed in the section background (sub-section Perceptual Facts), followed by a similar review at the PhL sub-section Neuro-Physiological Facts). This will be followed by the proposition of a bottom-up methodology in section Neuromorphic Speech Processing where a Neuromorphic Hierarchical Architecture for Speech Processing (NHASP) is presented. Results from current simulations will be presented and discussed in the section Some Selected Results. Future research lines will also be reviewed (section Future Research Directions). The section conclusions will close the work presenting the most relevant reflections derived from the research discussed.

BACKGROUND

The issue of bio-inspired or neuromorphic speech processing has been a matter of discussion since audition started to be studied scientifically (Helmholz, 1885). But it is since mid-twentieth century when the advancement in physiological and psychological studies opened a new door to a set of applications, such as hearing aids and prostheses, which had to rely in a serious systemic comprehension of the processes involved, requiring the formulation and testing of new systemic paradigms (Gold & Pumphrey, 1948, Gold, 1948, Flanagan, 1960a, Flanagan, 1960b). The interest in cochlear modeling (Allen, 1985) produced results as gammatone filters (Katsiamis, 2007) which were directly inspired in the time-frequency filtering capabilities of the Peripheral Auditory System (PAS). Many works were conducted to infer if these methods could be applied in general speech processing (Ghitza, 1988). Part of this research was also applied to the construction of integrated circuits reproducing the behavior of the PAS for intra-cochlear stimulation (López-Poveda, 2001). The proposition of using bio-inspired knowledge in speech processing came as a consequence, which raised immediate answers both in support and in opposition to the underlying idea. This controversy was neatly expressed by Hynek Hermansky (1997) in his famous sentence “Should speech recognizers have ears?” The answer to this important question is given by Hermansky himself: speech perception is very robust to noise, channel distortions, or multiple source effects. Besides, he claimed that bio-inspiration was already present in speech processing in Bark and Mel Scales, equi-loudness curves, Perceptual Linear Predic-