Vowel phonologic symbol assignment by neuromorphic speech processing

Pedro Gómez, Cristina Muñoz, Victoria Rodellar, Rafael Martinez
Grupo de Informática Aplicada al Procesado de Señal e Imagen (GIAPSI), Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid
e-mail: pedro@pino.datsi.fi.upm.es, tel.: +34.91.3367384, fax: +34.91.3366601

Abstract:

Recent research in the physiology and functionality of the auditory processing pathways is helping to unearth the ways in which the human brain processes speech and language [1, 2]. This knowledge is of great help to understand the role played by auditory signal perception from the phonetic-acoustic pattern distribution space to the phonologic representation space. Based on these concepts the present work introduces some of the basic facts already well established to explore plausible procedures which may be used by the natural language processor of humans in constructing vowel phonologic spaces and in using these to discriminate and assign phonologic values to perceived acoustic patterns. These hypothetical procedures are tested with real speech examples by means of a Neuromorphic Speech Processing Architecture [3, 4] which is built using simple units resembling true biophysical neuronal structures under the functional point of view (Fig.2 left). The most relevant neuronal circuits intervening in auditory acoustic processing and in vowel space representation will be presented and discussed. Examples for a reduced vowel phonologic space as that of Castilian Spanish are given (see Fig.1). As a summary Fig.2 (left) shows the results of automatic subdivision of the reference sentence into static and dynamic segments (top) and the assignment of phonologic representation (bottom, codified in colors, see figure caption). A detailed presentation of the inner neuron membrane activation levels and firing sequences will be presented as part of the results, and the adaptive and statistical procedures for dynamic tracking, mutual exclusion and symbol assignment will be commented. Possible applications to Automatic Speech Recognition, Phonetic Boundary Detection and Labeling, Phonologic Symbol Spotting, Speaker Recognition and related problems will also be addressed [5, 6].

References:

Figures cited in the abstract:

Fig. 1. Top: time series of the utterance -es hābil un solo día- (esβIoʊnsIoʊdIæ) from a male speaker. Middle: Adaptive Lineal Prediction Spectrogram (grey background) and first two formants (superimposed in color). The color dots mark the positions of each pair \((F_1,F_2)\) from green (the oldest) to red (the most recent). An approximate phonetic labeling is given as a reference. Bottom Left: Formant plot of \(F_2\) vs \(F_1\). Bottom Right: Same plot as a Formant Chart commonly used in Linguistics. The black circles give the vowel triangle center of gravity and the centroids of its extremes. The blue triangle and circles give the limit positions of the five reference vowels \(\{/i/, /e/, /a/, /o/, /u/\}\) (male speaker in blue, female in melba). These plots show the formant trajectories of the utterance. There is color correspondence between the bottom and middle templates to track formant trajectories on the time axis.

Fig. 2. Right Template: Vowel processing and representation sections of the Neuromorphic Speech Processing Architecture described in [3, 4]. Upper data-flow pipe-line: Spectrogram Estimation Front-End, Lateral Inhibition Formant Profiling, Tonotopic Band Tracking, Vowel Band Grouping, Vowel Assignment by Mutual Exclusion and Vowel Temporal Clipping. Lower data-flow pipeline: Static Formant Tracking and Temporal Static Masking (see text for a detailed description). Left Template: Automatic vowel assignment to the reference set representation space \(\{/i/, /e/, /a/, /o/, /u/\}\). Top: Static-dynamic framing. Bottom: Phonologic Assignment (/i/-blue, /e/-cyan, /a/-green, /o/-yellow, /u/-red).