Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 81 (2016) 101 – 106

Procedia Computer Science

5th Workshop on Spoken Language Technology for Under-resourced Languages, SLTU 2016, 9-12 May 2016, Yogyakarta, Indonesia

Automatic Syllable Segmentation using Broad Phonetic Class Information

Bogdan Ludusan*, Emmanuel Dupoux

Laboratoire de Sciences Cognitives et Psycholinguistique

EHESS / École Normale Supérieure, PSL Research University / CNRS, Paris, France

Abstract

We propose in this paper a method for syllable segmentation based on the Sonority Sequencing Principle, principle by which the sonority inside a syllable increases from its boundaries towards the syllabic nucleus. The sonority function employed was derived from the posterior probabilities of a broad phonetic class recognizer, trained with data coming from an open-source corpus of English stories. We tested our approach on English, Spanish and Catalan and compared the results obtained to those given by an energy-based system. The proposed method outperformed the energy-based system on all three languages, showing a good generalizability to the two unseen languages. We conclude with a discussion of the implications of this work for under-resourced languages.

Peer-review under responsibility of the Organizing Committee of SLTU 2016

Keywords: syllable segmentation; sonority; broad phonetic class; posterior probabilities

1. Introduction

The syllable is the smallest prosodic units and it plays an important role in the description of all prosodic phenomena. Similarly to other speech annotations, syllable segmentation is a time consuming task and automation of this process is desired, in order to be able to process large datasets. Information about syllables is useful not only for phonetic analysis of corpora, but also in speech technology applications like speech rate estimation [1], or the automatic detection of prosodic events (e.g. acoustic prominence [2], prosodic boundaries [3]).

A popular automatic syllable segmentation method is based on the energy of the speech signal (e.g. [4,5]). It offers the advantage of being language-independent, but requires the setting of a number of parameters, and its performance is sensitive to recording conditions. Another approach for language-independent automatic syllable segmentation can employ knowledge from the phonological theory. In linguistics, sounds can be grouped in classes, based on various criteria. One such criterion is the manner of articulation and the division of the phonetic space based on this criterion

* Corresponding author.
E-mail address: bogdan.ludusan@ens.fr

Peer-review under responsibility of the Organizing Committee of SLTU 2016
doi:10.1016/j.procs.2016.04.036

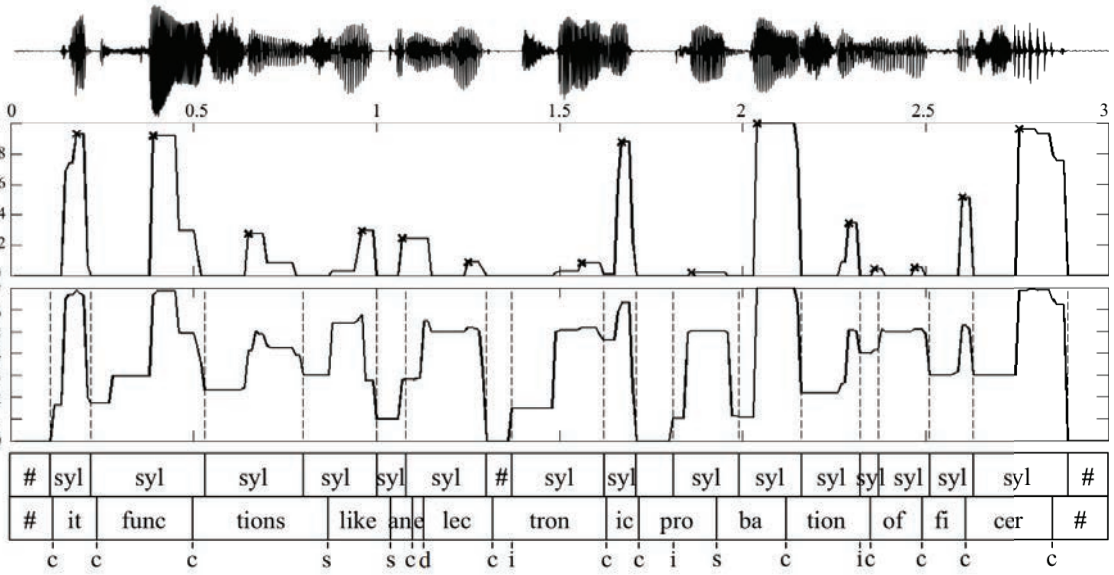

Fig. 1. Waveform of the phrase "It functions like an electronic probation officer." (upper panel) and corresponding nucleus sonority (middle panel), total sonority (lower panel) and its evaluation (bottom tiers). The position of the found nuclei is marked with an X sign in the middle panel, while the obtained syllable boundaries are marked by a dashed line in the lower panel. The upper tier contains the reference syllable segmentation, the middle tier the automatic segmentation (syl represents a syllable, while # a silence) and the lower tier the corresponding evaluation (c=correct, s=substitution, d=deletion, i=insertion).

will be called throughout the paper as broad phonetic classes. Each broad phonetic class has a different level of sonority, from obstruents, with a low sonority, to vowels, represented by a high sonority. For segmentation, one can apply the Sonority Sequencing Principle (SSP)[6], which states that the sonority inside a syllable increases towards the nucleus and and then decreases again towards the left edge.

We propose a system based on the SSP, which uses a speech recognizer, trained on an open-source corpus of English, to obtain the probabilities of each broad phonetic class. These probabilities are then combined with the sonority values of each class to derive an overall sonority function and syllable nuclei and boundaries are placed in correspondence to the maxima and minima of this function. Similar methods have been proposed for speech-based nucleus detection[7] and syllable segmentation[8]. A broad phonetic class recognizer was used to obtain the vocalic nuclei of syllable in order to estimate the speech rate[7]. Automatic syllable segmentation was performed in[8] by force aligning the speech signal, then taking the sonority values of the obtained phonemes and placing syllable boundaries in correspondence to the minima of this function. Differently from these approaches, we do not use the recognizer to produce a sequence of phonemes/phonetic classes, but to determine the posterior probability of each frame and we derive from it a continuous sonority function. Thus, we are not limited only to the class decision taken by the recognizer[7], but can take into account the contribution of all the classes. Also, by using phonetic recognition, not forced alignment[8], we can apply it to languages that do not have trained acoustic models.

The paper is further structured: Section 2 presents in detail the two components of the syllable segmentation system, namely the speech recognizer and the nuclei and boundary placement function. The datsets used in the experiments and the results obtained are detailed in Section 3. The paper concludes with a discussion on the performance of the system and its possible use for under-resourced languages.

2. Methods

The segmentation procedure is performed in two steps: First, a speech recognizer is used to decode the input sequence into broad phonetic classes. Second, we use the posterior probabilities given by the recognizer to derive two