REVIEW: PAVEL MACHAČ AND RADEK SKARNITZL (2009). PRINCIPLES OF PHONETIC SEGMENTATION. PRAHA: NAKLADATELSTVÍ EPOCHA.

HYNEK BOŘIL

Knowledge of phone boundaries has been traditionally of interest in a variety of speech-related research disciplines such as speech pathology, linguistics, and cognitive science. Currently, with the growing effort to develop interfaces that would allow for more friendly and for us, humans, more natural communication with machines, also speech engineers, who have heavily favored the so-called ignorance modeling for the last four decades, begin to realize that incorporating a finer knowledge of the underlying structure of speech, and in particular, conducting a highly accurate segmentation of speech into phonetic units during the system design, will considerably benefit the quality of speech synthesis and automatic speech recognition. So what is the catch? When closely inspecting natural speech, it is at times difficult or even impossible to find unambiguous phone boundaries, since those, at least in a traditional sense, may not really exist. This is due to the transient effects of coarticulation when the place and manner of articulation of one phone continuously transforms to another. The purpose of this book is to study the effects of articulation and coarticulation and establish comprehensive guidelines to replicable manual phonetic segmentation of the speech signal.

In the introduction, the authors explain the motivation behind the use of speech segmentation in the phonetic research, followed by a critical overview of the existing labeling approaches. The reader is gradually introduced to the peculiarities of determining phone boundaries in natural speech and the inherent ambiguity of the task due to the complex mechanisms of speech production. Subsequently, the authors stipulate a framework comprising a set of segmentation rules and methodology of their application. The framework is designed with an effort to alleviate uncertainties in the labeling process and provide consistent output across human labelers. The rest of the book provides particular recipes for boundary placement in individual phone classes and their sequences. Plosives, fricatives, nasals, trills, glides, laterals, and sequences of two obstruents, obstruent–liquids, consonant–consonant with the same manner of articulation, and plosive-like glottal stops and creaky glottal stops are covered in individual chapters. The authors analyze each phone class from the articulatory and acoustic perspective and search for a set of class-unique cues. These cues are then used to define class-dependent segmentation rules. Each chapter is

accompanied with a number of waveform and spectrogram plots that help the reader visualize the rules and become confident in their application. The real impact of the novel methodology is demonstrated in the final chapter, where the authors apply the proposed set of segmentation rules to the portion of a real-world speech corpus, reaching a superior segmentation consistency across human labelers compared to performances found in the state-of-the-art literature.

The content is accessible to any reader with elementary knowledge of phonetics, acoustics, and signal processing. The text is characterized by clear, seamless wording that makes the newly introduced concepts easy to digest. The selection of the book chapters is intuitive and their breadth represents a well balanced mixture of generally accepted facts and experimental evidence leading to the definition of segmentation rules.

A minor comment to the organization of the text – one might argue that adding a brief summary of techniques used for inter-labeler agreement evaluation, such as defining and demonstrating the usage of the kappa coefficient, might also benefit the reader and make the text even more comprehensive. On the other hand, such information can be easily found in other literature.

It has to be noted that up to now, there has been a lack of generally accessible set of rules for phonetic segmentation. As a result, research labs, and often even individual labelers, have been given no other choice than to establish their own *ad hoc* segmentation rules. No matter whether such rules were resulting from a life-long research experience or were established on-the-fly to fit a particular application purpose, they were in most cases hard to follow by others – partly due to the limited page count allocated to research articles, and presumably partly due to the lack of interest in sharing such knowledge. However, the inconsistency in phonetic segmentation in literature violates the requirement of experiment replicability by others and in consequence, leads to discrepancy in the published research. For this reason, availability of a segmentation 'handbook' providing clear and unambiguous guidelines for determining phone label boundaries may have an invaluable impact on unifying the speech community's view on phone segmentation and will strongly benefit every individual who finds themselves in the need of phonetic corpus preparation.