Adaptive Speech Synthesis of Albanian Dialects

Michael Pucher\textsuperscript{1}\textsuperscript{(2)}, Valon Xhafa\textsuperscript{2}, Agni Dika\textsuperscript{2}, and Markus Toman\textsuperscript{1}

\textsuperscript{1} Telecommunications Research Center Vienna, Vienna, Austria
\{pucher,toman\}@ftw.at
http://www.ftw.at
\textsuperscript{2} Department of Computer Engineering, University of Prishtina, Prishtina, Kosova
\{valon.xhafa,agni.dika\}@uni-pr.edu
http://www.uni-pr.edu/

Abstract. In this paper, we show how adaptive modeling within the statistical parametric speech synthesis framework can be applied to Albanian dialects. We develop speaker dependent voices for the Tosk and Gheg dialect and adapt models for the Gheg dialect from the Tosk models. We show that the adapted Gheg models outperform the speaker dependent Gheg model on an intelligibility and dialect classification task. Furthermore we show that the speaker dependent Tosk model outperforms a formant based synthesizer on an intelligibility, dialect classification and pair-wise comparison task. This formant based synthesizer is the only publicly available synthesizer for Albanian at the moment. We also show that our Gheg and Tosk synthesizers are as intelligible as natural speech. The method where one dialect is modeled through adaptation of a closely related other dialect can be applied to language varieties in general, where the background variety and adapted variety can be chosen based on pragmatic considerations like speaker or data resource availability.

Keywords: Speech synthesis · Albanian · Adaptation · Dialect

1 Introduction

Adaptive parametric HMM-based speech synthesis \cite{1,2} allows for the usage of a background model or average model to improve synthetic voice quality with small amounts of adaptation data. The authors of \cite{3} applied the adaptive framework to the synthesis of Austrian German dialects. In this paper, we use an Albanian Tosk dialect background model to improve an Albanian Gheg dialect adapted voice.

The adaptive approach that works with small amounts of data is especially interesting for languages such as Albanian where not so many language resources are available. Today the only open-source available synthesizer for Albanian is a formant-based synthesizer that is still in a “provisional” development stage \cite{4}. The synthesizers developed for this study are based on open-source components \cite{4,5} and we plan to release open-source synthesizers for Albanian.
The method we use here can be used for any variety pairs of phonetically closely related varieties (dialects, sociolects, and accents). We can use the variety where it is easier to collect a larger amount of data to train a background model. Then we can adapt the variety where it is more difficult to obtain large amounts of data. It can be difficult to obtain data when speakers are difficult to find, or language resources like lexica, grapheme-to-phoneme rules, or texts for recording scripts are not available. In our case we choose the Tosk dialect, which is also the basis for Standard Albanian, as background model since it was easier to obtain the large amount of text data that is needed to select a recording script by solving the respective set-cover problem. Furthermore, we could use a slightly modified grapheme-to-phoneme conversion from an existing synthesizer for Tosk [4].

2 The Albanian Language

The Albanian language belongs to the family of Indo-European languages. In the tree of languages, the Albanian language does not share any descent connection with other member languages of this family and is presented as a separate branch that grows from the root of the tree. This classification is based on phonological, morphological and other features [6].

There are two basic dialect forms of the Albanian language: Gheg and Tosk [7,8]. Furthermore there are also mixtures between Gheg and Tosk dialects, like Arbërësh and Arvanitika [7]. In countries where Albanian is spoken the official language is Standard Albanian, based on the Tosk dialect [9]. It is used in institutions, newspapers, and books. Gheg is spoken in more informal settings. For these reasons we adapted Gheg from Tosk, using the Tosk language in our background model. The Tosk dialect has 7 vowels and 29 consonants. The Gheg dialect uses long and nasal vowels, which are absent in Tosk [7].

Albanian almost has a one-to-one correspondence between letters and phones, which make grapheme-to-phoneme conversion easier, compared to some other languages.

3 Grapheme-to-Phoneme Conversion

Grapheme-to-phoneme (G2P) rules for the Tosk dialect were taken from an existing speech synthesizer [4]. Some errors of the G2P rules had to be corrected for Tosk and additional rules had to be introduced for Gheg as shown in Table 1. In general the G2P problem is relatively simple to solve for Albanian since there is an almost one-to-one mapping between letters and phones.

The additional rules for Tosk and Gheg shown in Table 1 are context independent, i.e. the characters “dh” are replaced everywhere by “D”. Table 1 also shows the International Phonetic Alphabet (IPA) symbols in brackets. Due to phonological differences between Tosk and Gheg we need to introduce additional rules for Gheg. The Gheg dialect has the nasal vowels “È” and “À”. These nasal vowels are mapped to “nA” and “nE” in our phone set.