

Andrzej Sawicki<sup>1</sup>, Piotr Zubrycki<sup>1</sup>, Alexandr Petrovsky<sup>1</sup>

## DESIGN OF TEXT TO SPEECH SYNTHESIS SYSTEM BASED ON THE HARMONIC AND NOISE MODEL

**Abstract:** This is a proposal of concatenative text to speech synthesizer for the Polish language, based on diphones and "Harmonics and Noise Model" (HNM). HNM has been successfully applied on a speech encoder and decoder, resulting in a high-quality of processed speech at low bit rate. Applying this model to speech synthesis system allows obtaining good quality of synthesized speech, and the small size of database parameters.

The proposed project consists of two main modules. The Natural Language Processing (NLP) is used to analyse and convert the written text for phonemes and diphones using morphological rules. NLP discovers at the same time prosodic features for later modification of synthesized speech parameters in order to obtain the stress and voice intonation. The second section is a synthesis system, derived from speech decoder, preceded by a system of adapting the parameters of speech based on prosodic rules.

The system of speech synthesis from the parameters is working in the frequency domain and uses the frequency spectrum envelope, which easily allows modifying the frequency, amplitude and duration of the signal when applying the prosodic rules. The algorithm of continuous phase designation at the speech frame borders allows concatenating portions of synthesized speech and diphones without phase distortion on the merger. Speech synthesizer operates on the diphone database, created applying fragmentation of recorded speech signal representing the pairs of phonemes. Sounds related to diphones are analyzed by speech encoder. It provides the parameters that described harmonic and noise components of speech, using the linear prediction filter LSF coefficients, resulting in a small size of diphone database.

**Keywords:** Speech synthesis, TTS, harmonic and noise model

## KONCEPCJA UKŁADU SYNTEZY MOWY Z TEKSTU OPARTEGO NA MODELU HARMONICZNE I SZUM

**Streszczenie** Artykuł przedstawia projekt konkatencyjnego syntezy mowy z tekstu dla języka polskiego, opartego na difonach i modelu Harmoniczne i Szum. Model Harmoniczne i Szum został z powodzeniem zastosowany w układzie kodera i dekodera mowy, dając

---

<sup>1</sup> Faculty of Computer Science, Białystok Technical University, Białystok

w rezultacie dobrą jakość przetwarzanej mowy przy niskiej przepływności bitowej. Zastosowanie tego modelu do układu syntezy mowy pozwala na uzyskanie dobrej jakości syntezy mowy, oraz niewielki rozmiar bazy parametrów.

Układ składa się z dwóch głównych modułów. Moduł Naturalnego Przetwarzania Języka służy do analizy i zamiany tekstu pisanego na fonemy oraz difony, przy wykorzystaniu reguł morfologicznych. Procesor tekstu wyznacza jednocześnie warunki prozodii związane z późniejszą modyfikacją parametrów syntezywanego głosu w celu uzyskania akcentowania i intonacji. Drugim układem jest moduł syntezy, oparty na dekoderyze mowy poprzedzonym systemem adaptacji parametrów mowy w oparciu o wyznaczone wcześniej reguły prozodyczne.

Układ syntezy mowy z parametrów działa w dziedzinie częstotliwości i bazuje na obwiedni spektrum, co w prosty sposób pozwala na modyfikację częstotliwości, amplitudy i czasu trwania sygnału przy stosowaniu reguł prozodycznych. Algorytm wyznaczania ciągłej fazy na granicach ramek sygnału mowy pozwala na łączenie fragmentów syntezywanego głosu oraz poszczególnych difonów bez zniekształceń fazowych na połączeniu.

Syntezytor mowy operuje na bazie difonów, stworzonej na podstawie fragmentaryzacji nagranych sygnałów mowy na części, reprezentujące połączenia par fonemów. Dźwięki odpowiadające difonom są analizowane przez moduł analizy mowy. Dostarcza on ciąg parametrów reprezentujących harmoniczne i szumowe komponenty sygnału mowy, opisane za pomocą filtrów liniowej predykcji i współczynników LSF, dając w rezultacie niewielkiej wielkości bazę difonów.

**Słowa kluczowe:** Synteza mowy, model harmoniczne i szum

Artykuł zrealizowano w ramach pracy badawczej W/WI/6/09.