

ARTIFICIAL INTELLIGENCE LABORATORY (AILAB)
INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF LATVIA

Latvian Text-to-Speech Synthesizer

Mārcis Pinnis

Marcis.Pinnis@lumii.lv

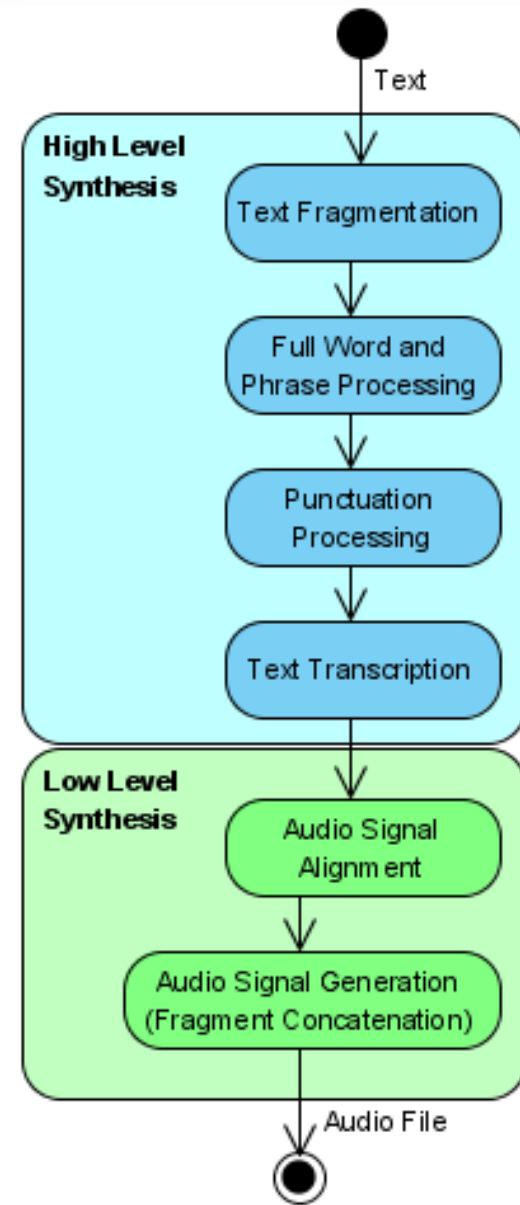
Ilze Auziņa

Ilze.Auzina@lumii.lv



Approach and Features

- AILAB IMCS UL Text-to-Speech system T2S V1
 - Concatenative text-to-speech system
 - The system features:
 - variable length speech fragment concatenation
 - diphones
 - full words
 - common phrases
 - multiple sound combination fragments
 - Punctuation and silence fragment length control
 - Rule based text transcription process (in order to obtain the phonetic representation of a text)
 - Audio fragment concatenation with interpolation at signal concatenation points to force signal smoothing



T2S V1 - Domain Oriented System

- The flexible speech fragment length allows domain orientation to achieve better synthesis results
 - T2S V1 domain oriented for Weather Forecasts



Issues in Development

- Several Issues arose in the Development of the T2S V1 Speech Synthesis System
 - Orthographic ambiguities in characters
 - “e” - /e/ “egle”, /{/ “ezers”
 - “ē” - /e:/ “ēvele”, /{:/ “ēka”
 - “o” - /uo/ “ola”, /o/ “omlete”, /o:/ “oda”
 - Sound segment alignment isn’t always smooth
 - Synthesized speech is too neutral – prosody is not modeled
 - System’s current speed is not suitable for on-the-fly applications



Unsolvable Issues

- The Latvian Language orthography allows the usage of “e” and “o” for more than one phoneme, which makes it impossible to guess the right pronunciation.
 - “ēdu” – is it present or past?
 - “koks” – is it a microorganism or a tree?
 - “deva” – is it a noun or a verb?
- Such issues can be solved only if the context is large enough to guess the right form. If the context is not present (Consider the sentence “Es ēdu pusdienas.”) or is not wide enough, prediction is theoretically impossible.



Demonstration

Konkatenatīva latviešu valodas runas sintēzes sistēma

Teksts:

Sintezēt

1. Tuvākajās dienās saglabāsies mākoņains laiks, tomēr brīžiem skaidrosies.
2. Smidzina lietus. No rīta mākoņains. Vēss. +8...+5 °C. vējš 4 m/s
3. Valsts prezidenta Valda Zatlera svētku uzruna.



Atskaņot



The Perspective of Further Research

- The system may be improved in three ways:
 - By introducing better NLP solutions:
 - Context dependent abbreviation analysis
 - Context dependent numeric transformation analysis
 - Context dependent morphological analysis
 - Sentence and word level prosody analysis
 - Phonetic dictionary necessary to minimize the impact of wrong rule application
 - By introducing better low level synthesis:
 - Usage of PSOLA and RELP approaches for prosody control
 - Alternative – switch to HMM-based unit selection speech synthesis (for instance, HTS)
 - Algorithm optimization (solves the speed issue)
 - By introducing higher quality speech corpus:
 - Better target domain vocabulary coverage
 - Better speech fragment alignment



THANK YOU 😊



QUESTIONS?