ICSP 2008

Towards Automatic Audio Track Generation for Czech TV Broadcasting:

Initial Experiments with Subtitles-to-Speech Synthesis

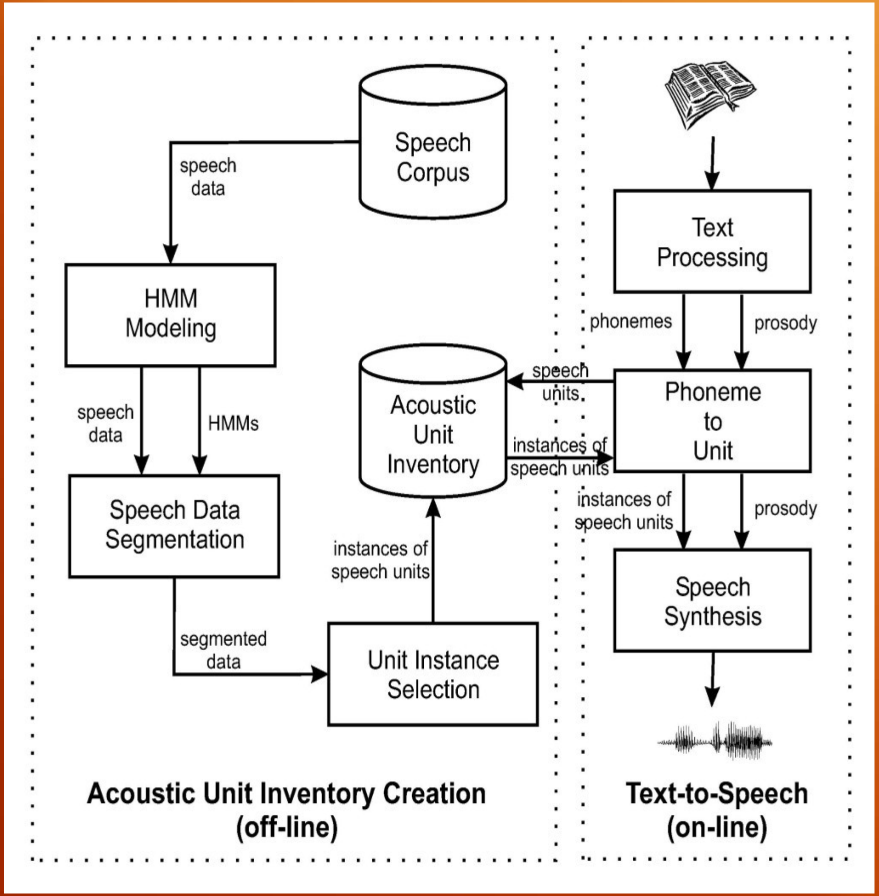
Zdeněk Hanzlíček, Jindřich Matoušek and Daniel Tihelka
University of West Bohemia, Czech Republic

Introduction

- project ***Elimination of the Language Barriers Faced by the Handicapped Watchers of the Czech Television*** – 2 main objectives
    - real-time subtitling (speech recognition)
    - automatic generation of audio track from subtitles (speech synthesis)
- subtitles (closed captions)
    - broadcasted by using teletext page no. 888
    - EBU subtitling data exchange format
        - text + timing
        - no speaker information

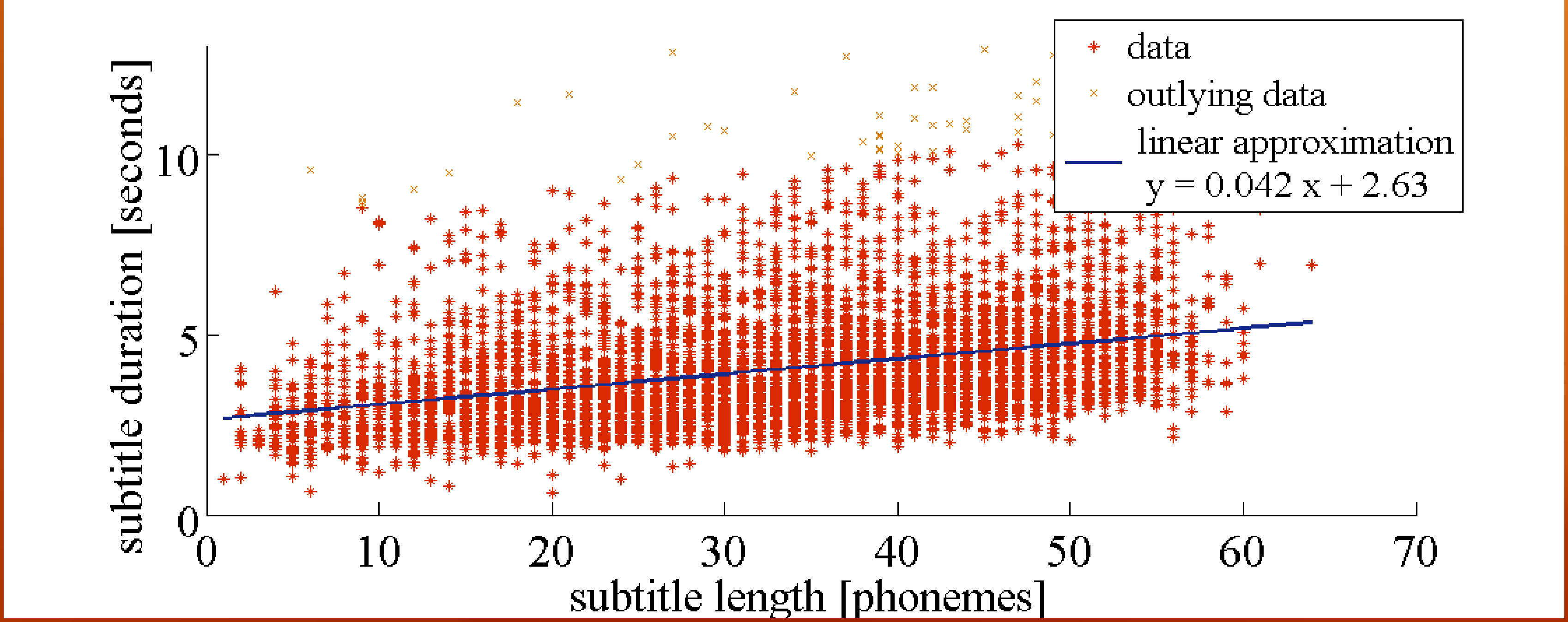
TTS system ARTIC

- ARTIC = artifical talker in Czech
- corpus-based concatenative speech synthesis
- 2 versions
  - single unit instance system
  - multiple unit instance system (unit selection method)
- several different voices (males and females)



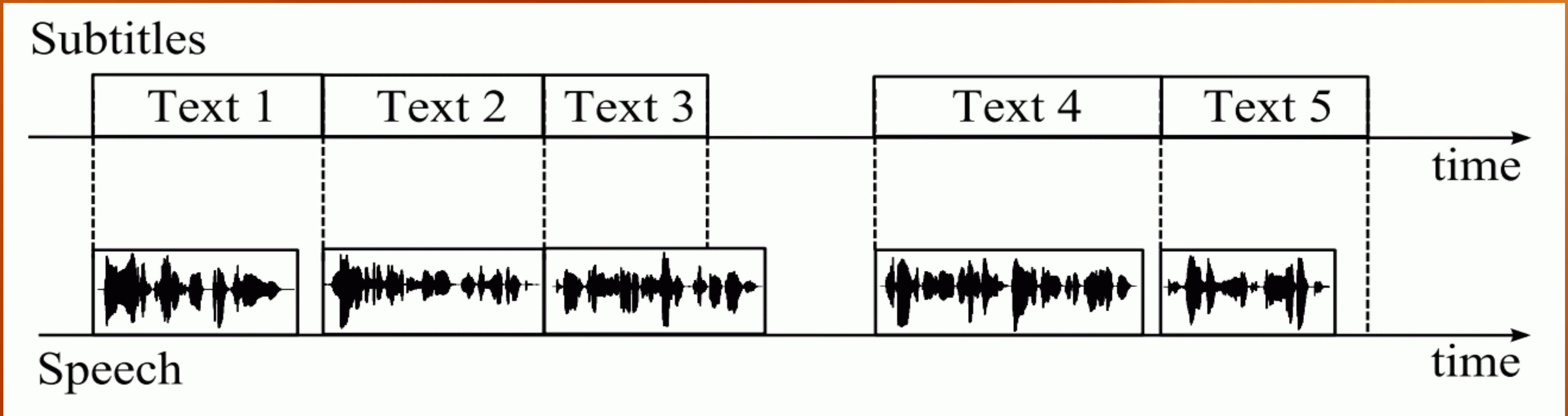
Subtitle analysis (1)

- subtitles for 20 various programmes (documentaries, talk-shows, cartoons, movies...)
- 5794 subtitles in sum



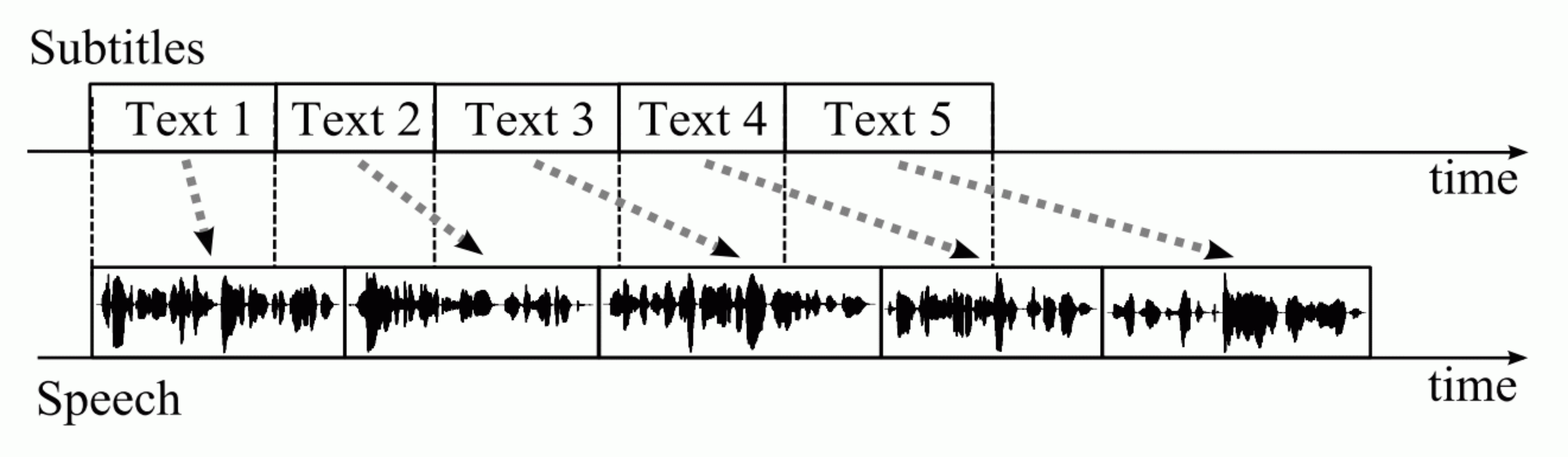
Timing desynchronisation (1)

- ideal case for subtitle-to-speech synthesis
  - no utterance overlaps into following subtitle time slot



Timing desynchronisation (2)

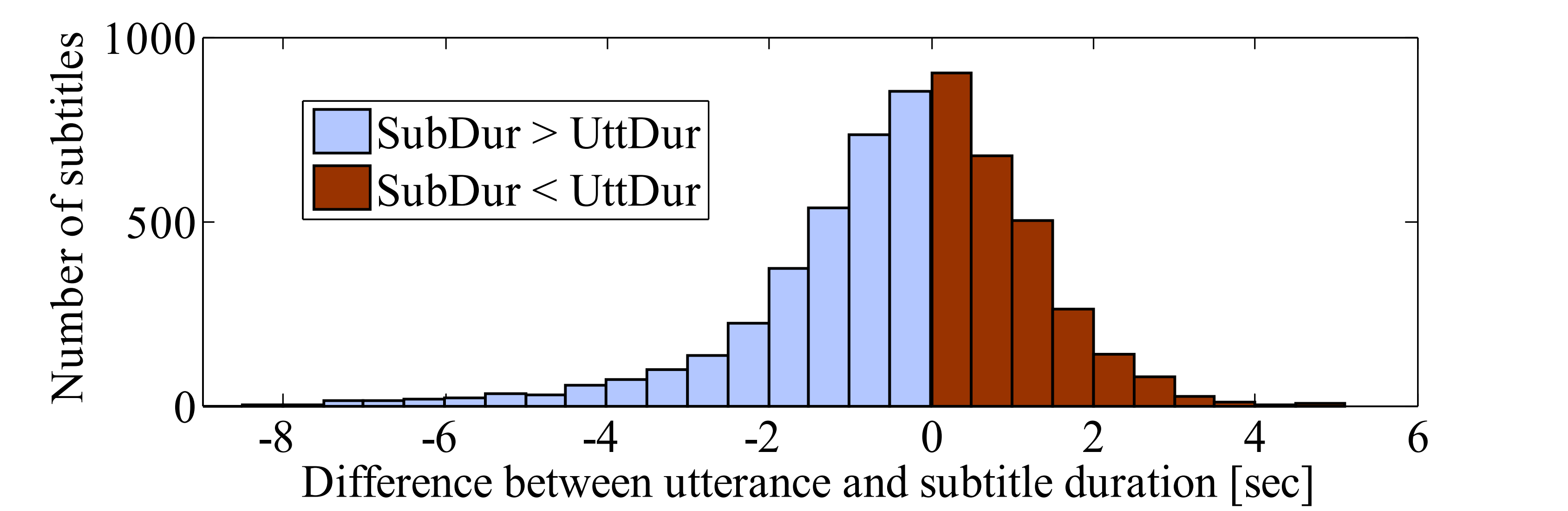
- serious problem
  - utterance overlaps into following subtitle time slot
  - utterances are delayed



Subtitle analysis (2)

- subtitle time slot length vs. utterance duration

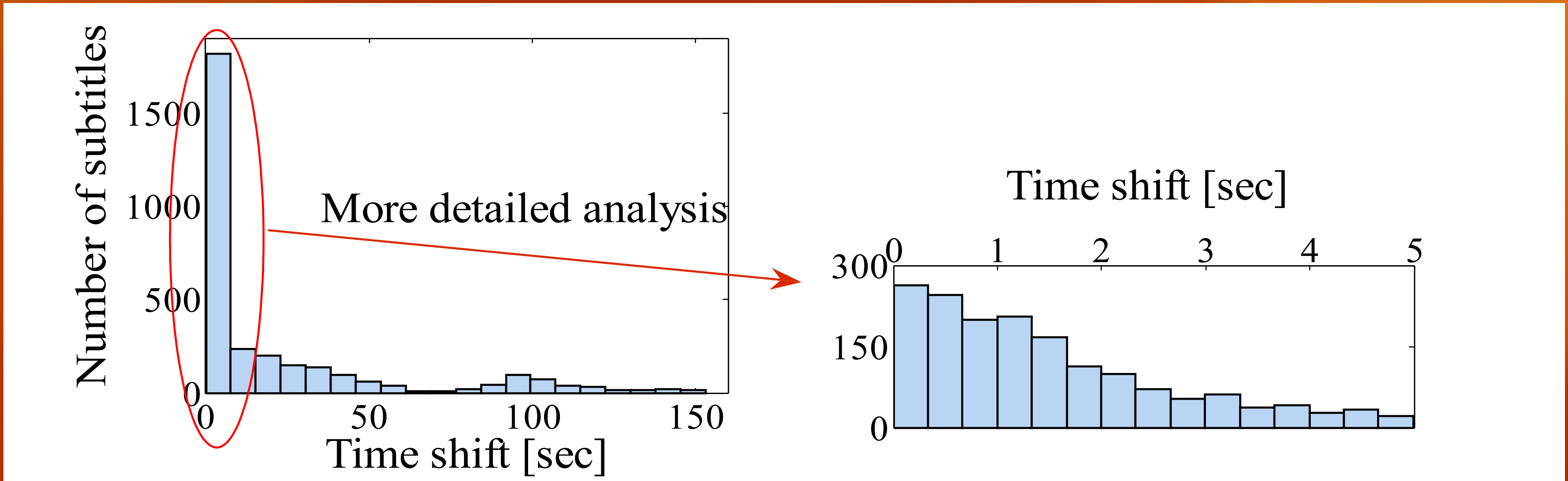
	VM	MM	MF	SM	SF
SubDur > UttDur [%]	61.2	53.3	51.9	77.1	86.3
SubDur < UttDur [%]	38.8	44.7	48.1	22.9	13.7



Subtitle analysis (3)

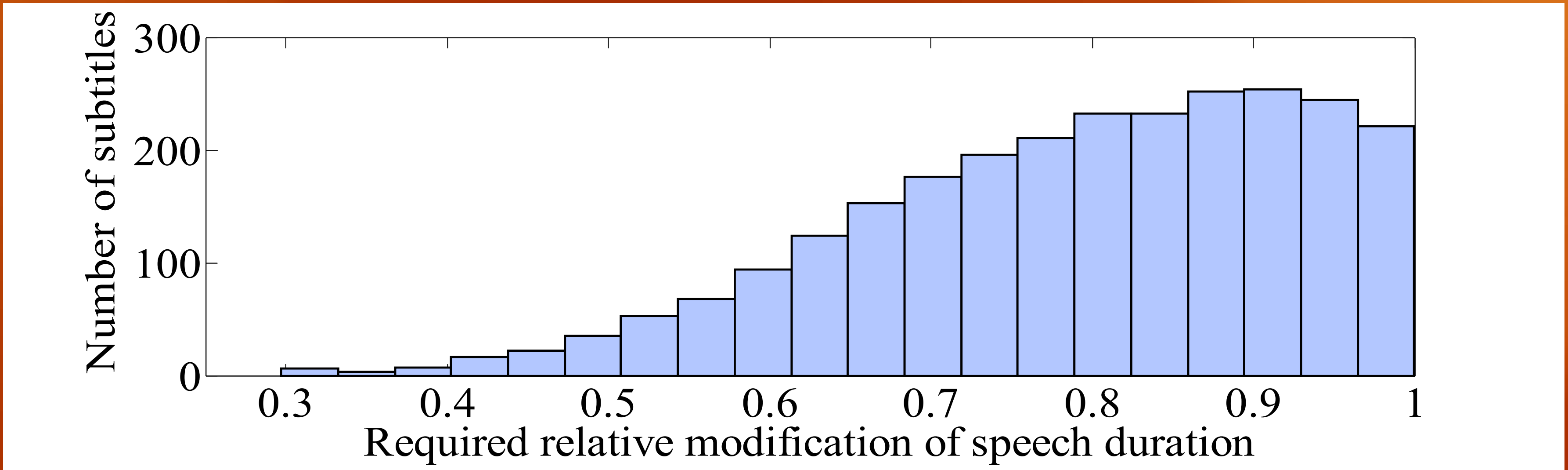
- utterance delay (time shift compared to subtitle display)

	VM	MM	MF	SM	SF
Correct begin [%]	54.6	39.4	35.7	72.3	84.7
Shifted begin [%]	45.4	60.6	64.3	27.7	15.3
Average delay [sec]	6.4	21.3	31.7	1.7	0.9



Problem solution

- faster speaker for curpus recording
- subtitle text abridgement
- selection of shorter speech units during synthesis
- time scale modification (WSOLA method)
  - speech corpus
  - synthesised utterances



Thank you for your attention.