

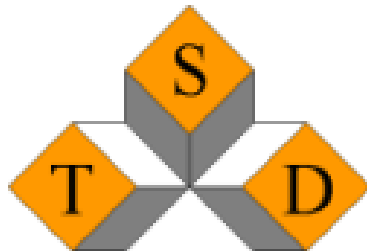
Special Speech Synthesis for Social Network Websites

Csaba Zainkó, Tamás Gábor Csapó,
Géza Németh

{zainko, csapot, nemeth}@tmit.bme.hu



BME TMIT, Hungary



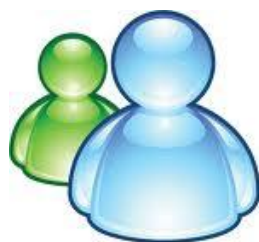
TSD Conference, Brno
September 10, 2010

Contents

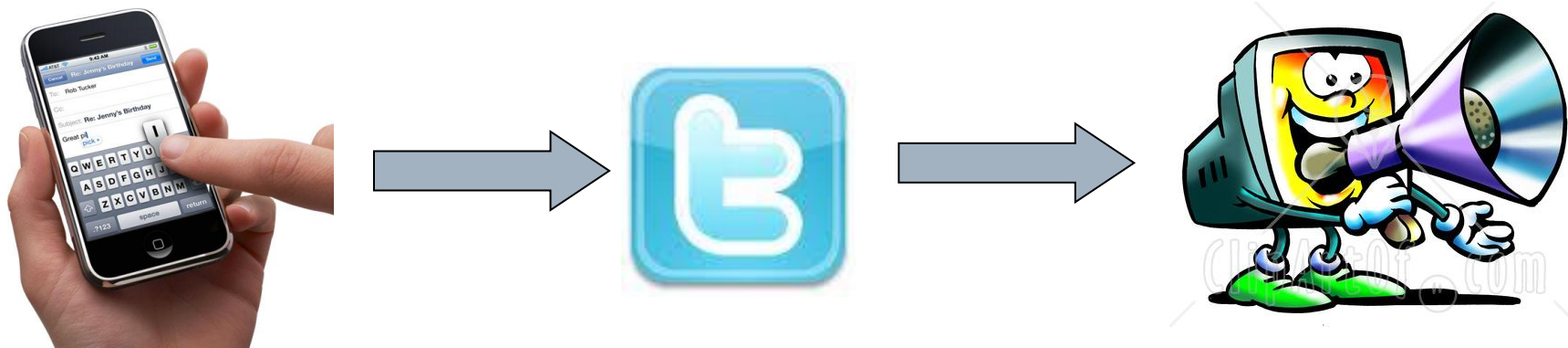
- Chat / microblog-reading
- Diacritics restoration
- Spontaneous-like speech
- Emotional synthesized speech
- Conclusions

Needs / goal of chat / microblog TTS reading

- ❑ Microblog websites (e.g. Twitter)
- ❑ Chat applications (e.g. MSN, Gtalk)
- ❑ Messages not too often
- ❑ Mobile environment



Chat / microblog-reading, plan



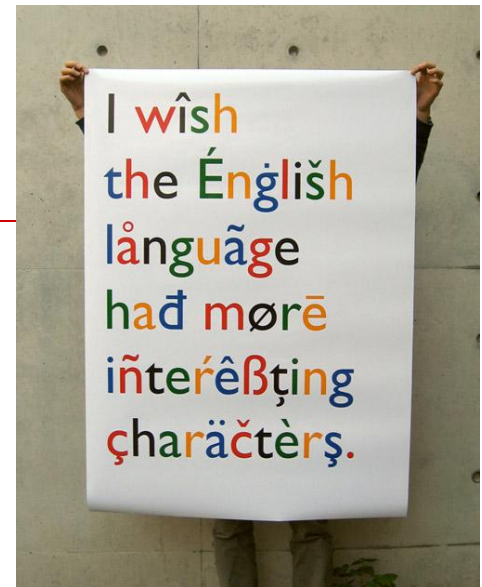
Problems of chat / microblog TTS reading

- Letters without diacritics
 - SMS, character encoding
 - Hard / slow to write diacritics (e.g. iPhone, iPad)
 - Diacritics restoration
- Emoticons
 - Spontaneous style
 - Emotional speech

Diacritics restoration /1

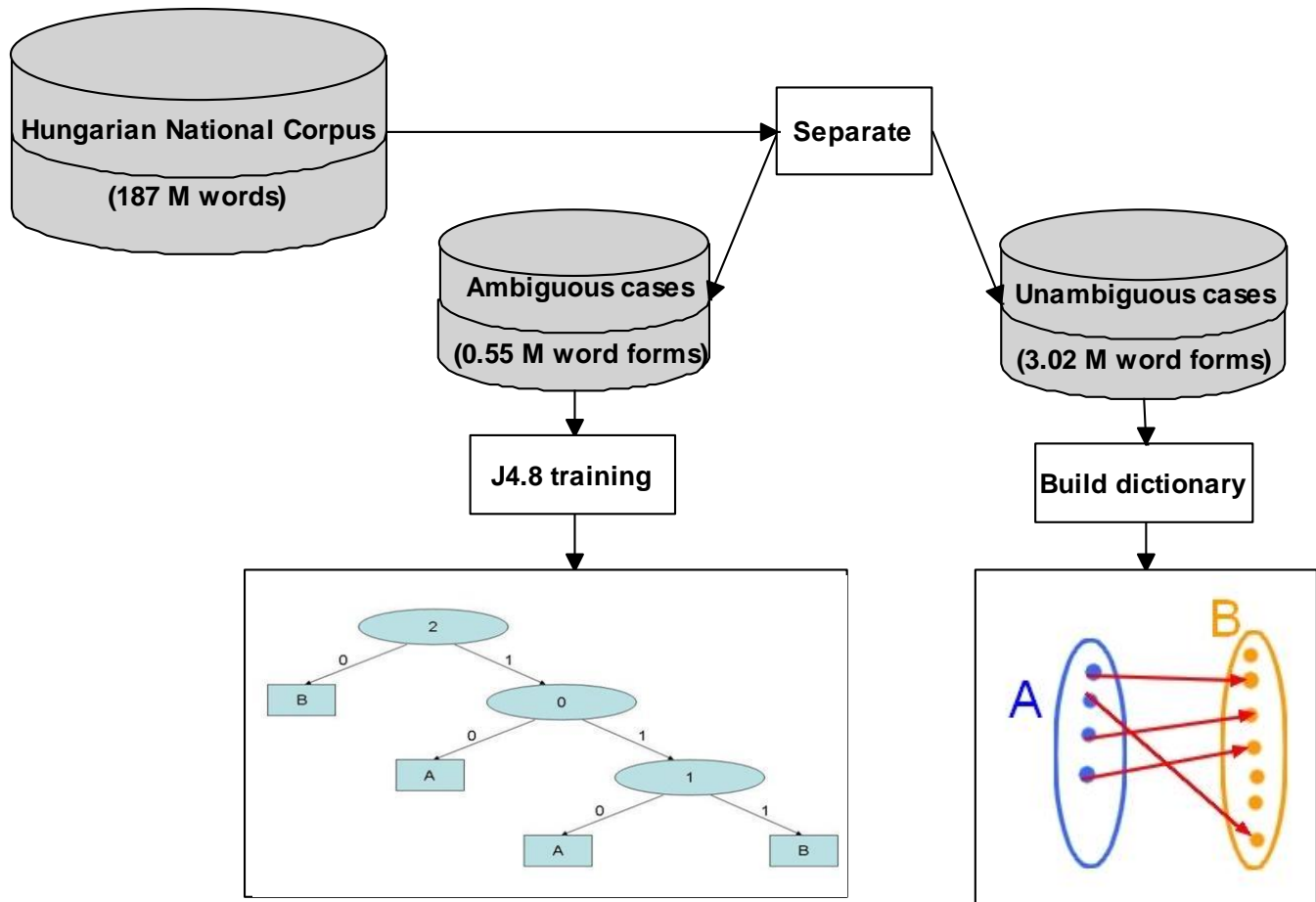
Intro

- Problem in most European languages
- Hungarian:
 - a-á, e-é, i-í, o-ó-ö-ő, u-ú-ü-ű
- Solutions for other languages
 - Dictionary-based (word probability)
 - HMM-based
 - Word level vs. Letter level



Diacritics restoration /2

Training



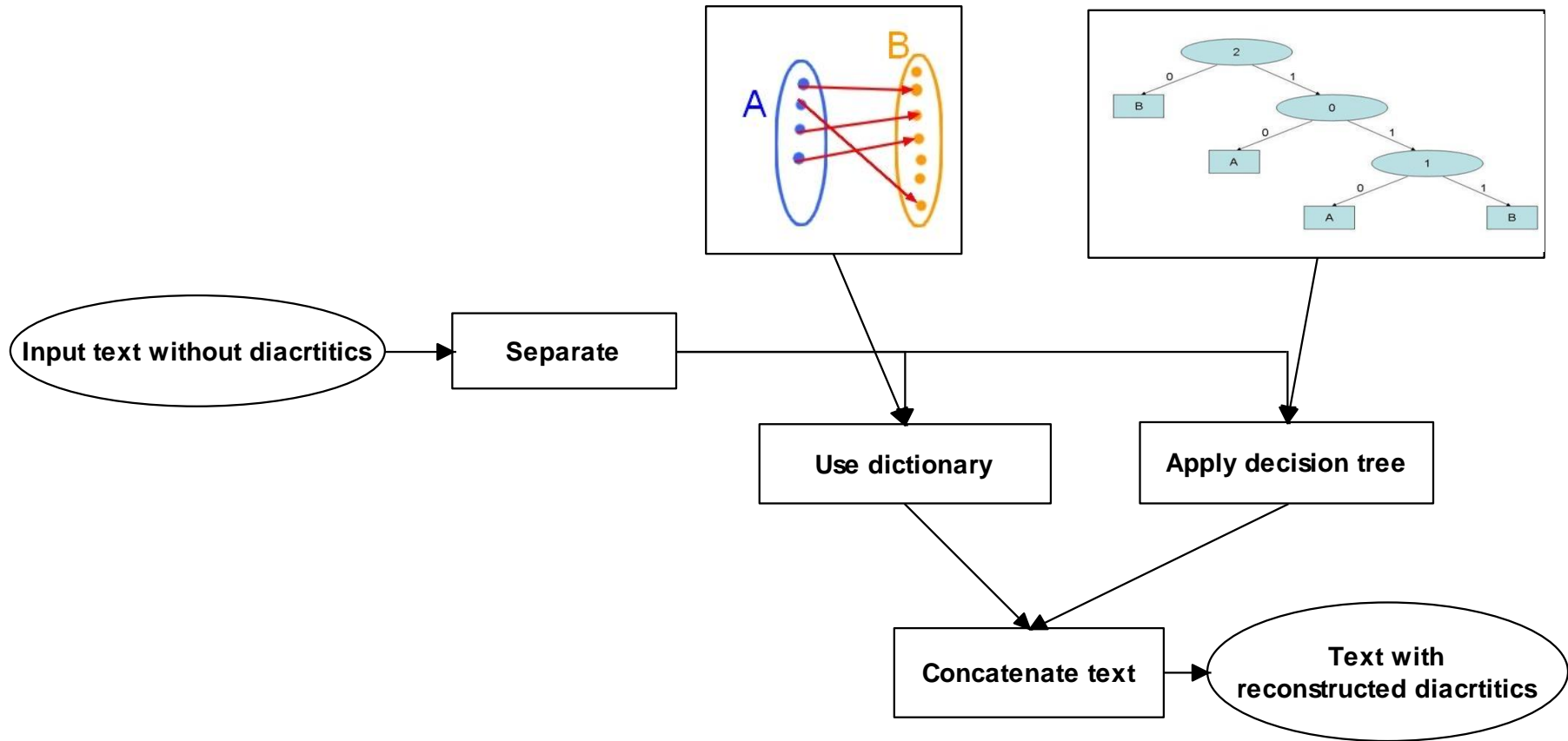
Diacritics restoration /3

Training

- HNC: 187 million words
 - 3.02 million unambiguous word forms (84.5%), e.g. „az” (the) ✓ „áz” ✗
 - dictionary
 - 0.55 million ambiguous word forms (15.5%), e.g. „meg” (plus) „még” (still)
 - J4.8 decision tree
 - 100 most frequent words separately
 - 20 letters context

Diacritics restoration /4

Use



Diacritics restoration /5

Accuracy

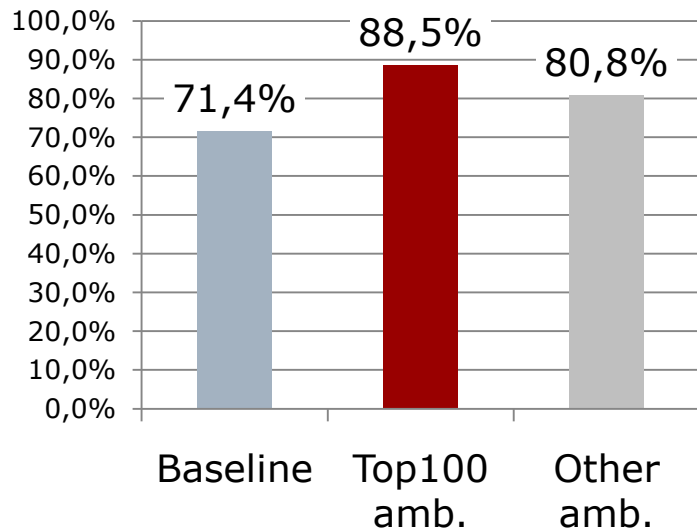
- All cases
(ambiguous + unambiguous)
- Word accuracies
 - 97.7% for „DIA” (Literature texts)
 - 97.2% for „Personal” (Web forum texts)
 - 98.2% for the whole HNC

Diacritics restoration /6

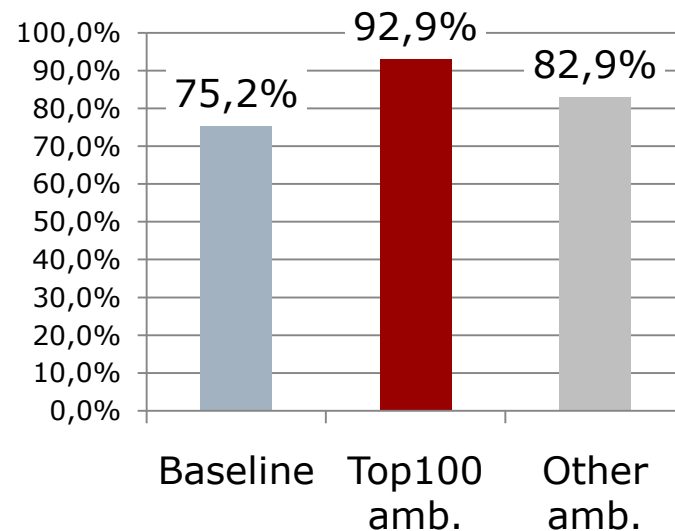
Accuracy

- Only ambiguous cases
- Word accuracies

Web forum texts



Hungarian National Corpus



Spontaneous speech

Differences compared to read speech

- Intonation contour and its variability
- Breaks, pauses (silent, filled: breathing)
- Less strict relation of prosody and syntax
- Disfluencies
- Lack of exact structure
- Redundancy, repetition of words
- Acoustic vowel reduction
- Final lengthening

Spontaneous-like synthesized speech /1 Method

- Corpus based TTS, read corpora
- Find conversational aspects
- Insertion of fillers (humming, hesitation, laughter)
 - After conjunctive words
- Insertion of breath
 - At phrase boundaries
 - In-phrase breath
- Pause timing
 - Variable pause lengths
 - More frequent pauses than in read speech



Spontaneous-like synthesized speech /2 Results, experiences

Output of corpus-based TTS

- Insert hesitation ✓
 - Disturbs understanding
 - Increased cognitive load
- Insert breaths ✓
 - Normal human function
 - Weak vs. Loud
- Insert laughter ✗
- Pause timing ✓
 - Acceptable if more frequent than in read speech

Emotional Synthesized Speech

/1 Intro

□ „emoticons” ⇒ emotions in speech

■ Neutral

■ Angry :@

■ Happy :)

■ Sad :(

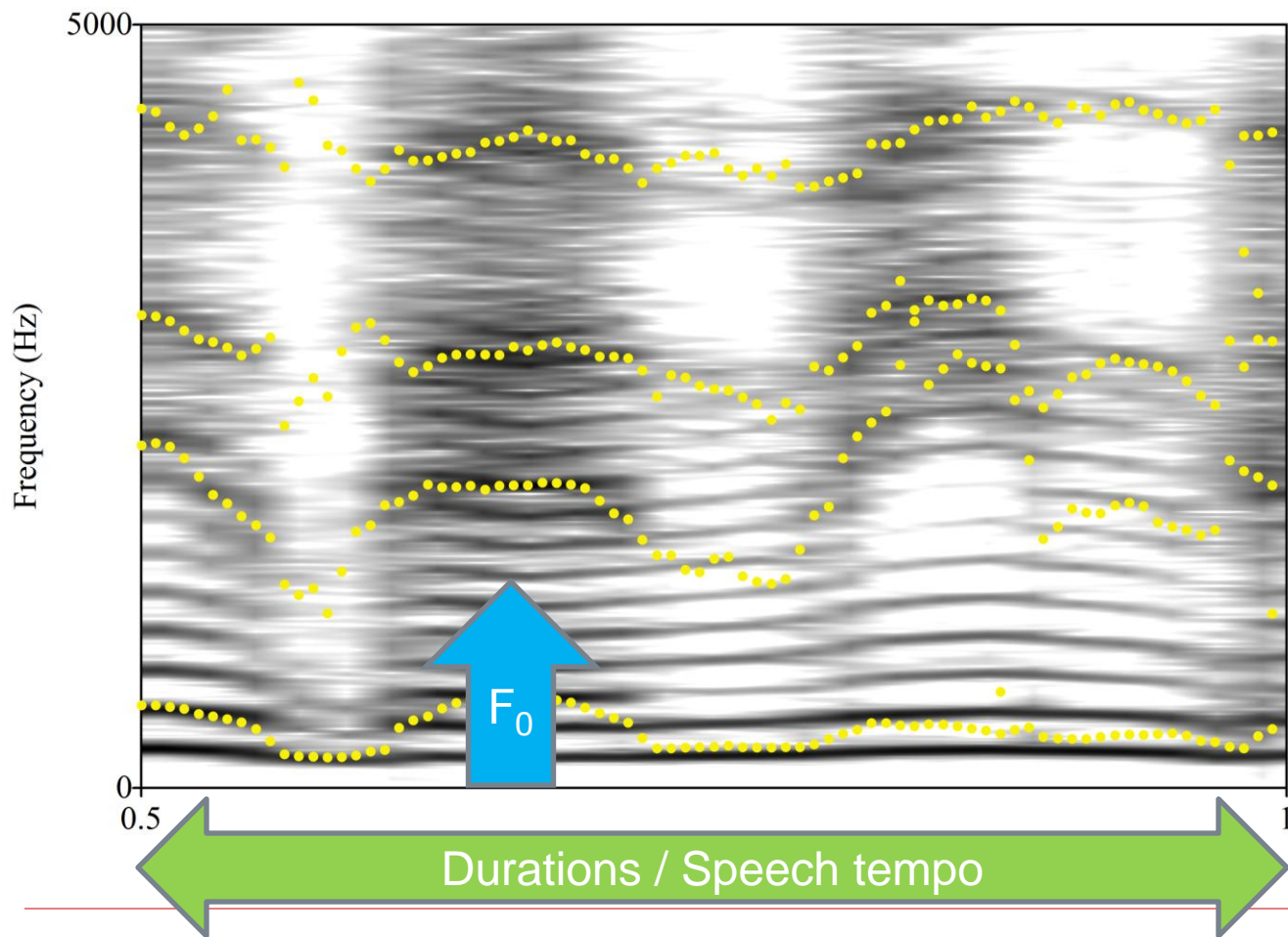


■ „A menüben minden szükséges információ elhangzik.”

■ (All necessary information is mentioned in the menu.)

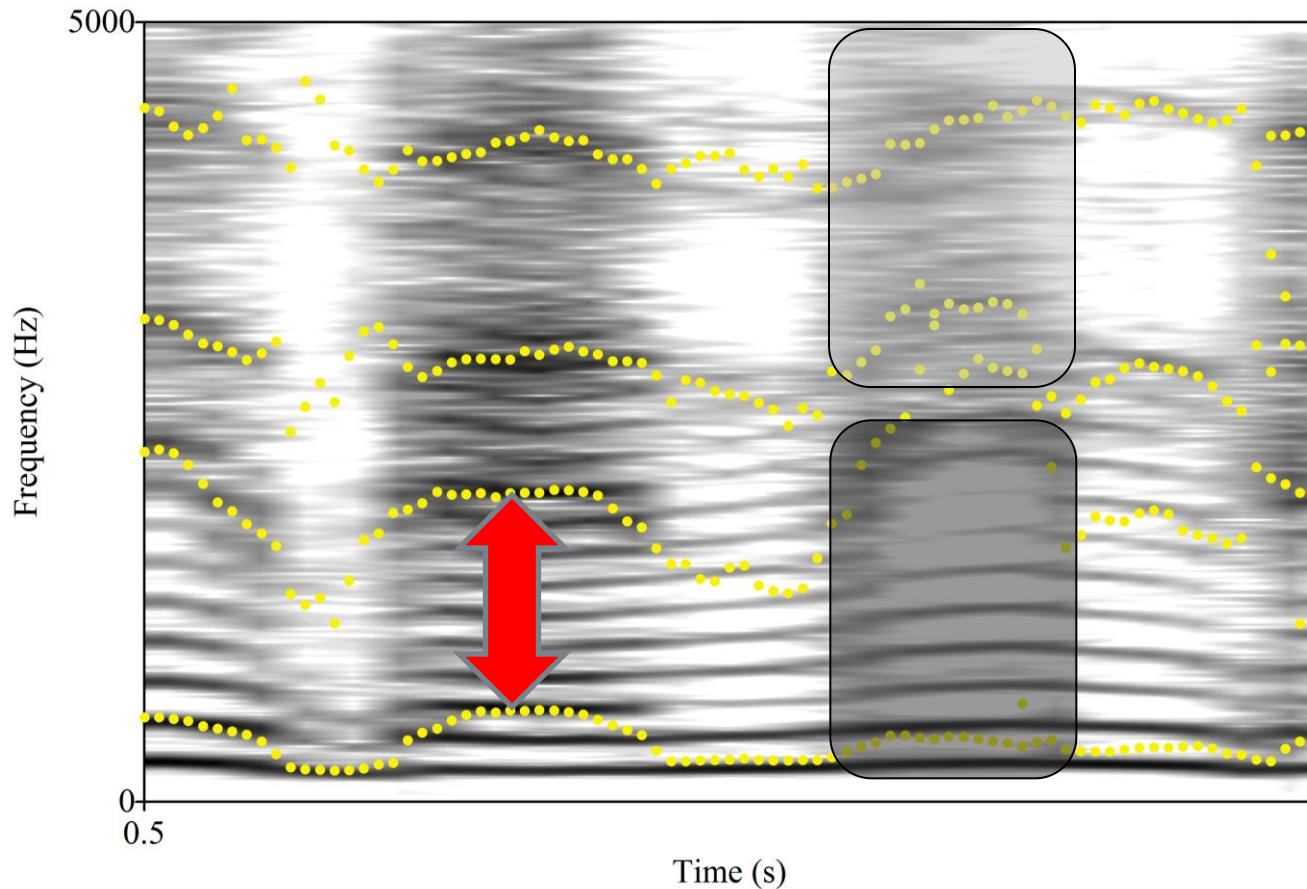
□ Forum texts: 90 thousand emoticons
in 1.5 million sentences

Emotional Synthesized Speech /2 Algorithm



based on
Přibilová
& Přibil
(2009)

Emotional Synthesized Speech /3 Algorithm



based on
Přibilová
& Přibil
(2009)

Emotional Synthesized Speech

/4 Experiment+Results

□ 3 sentences; Angry Sad Happy

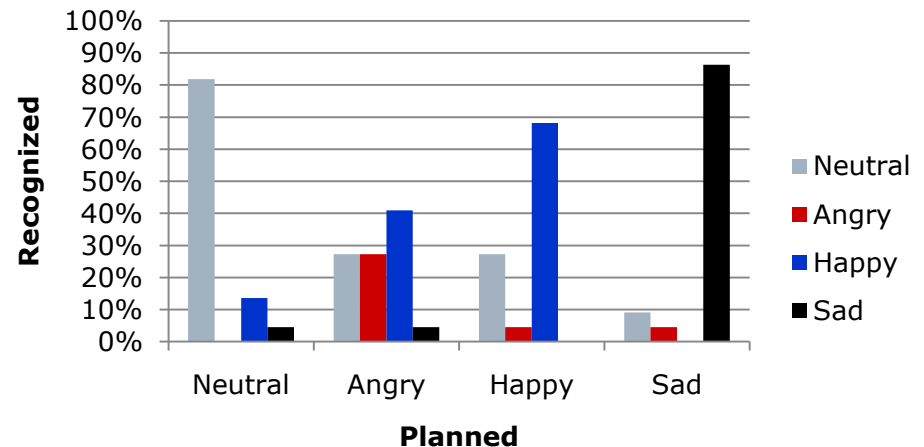


- „A menüben minden szükséges információ elhangzik.”
- (All necessary information is mentioned in the menu.)

□ 3 voices

- Natural speech
 - professional female
- TTS speech
 - female, male

TTS-female



Emotional Synthesized Speech

/5 Conclusions

Female ✓ (natural / TTS)

(this) Male TTS ✗

Neutral, happy, sad ✓

Angry ✗

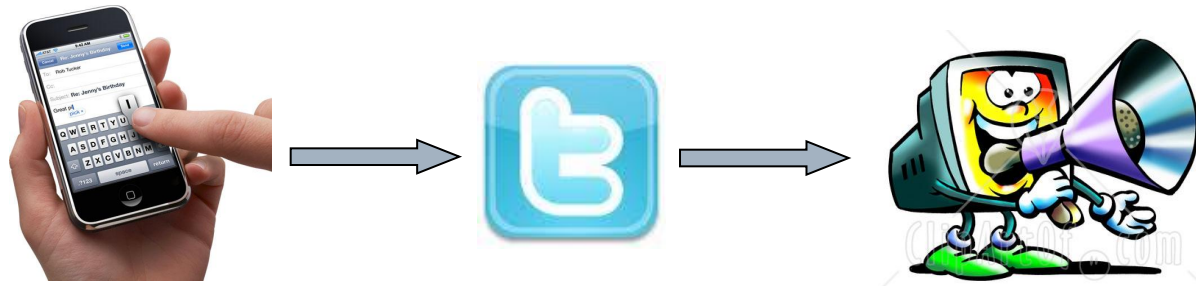
- confused with happy

- (human emotion recognition:

 - average 60-75%, Tóth et al. 2007; Scherer 2003)

Chat / microblog-reading, summary

- Diacritic restoration
- Spontaneous-like speech
- Emotional synthesized speech
- Other subproblems
 - language identification
 - spelling correction



□ csapot@tmit.bme.hu

- Acknowledgements: Supported by
- TELEAUTO project (OM-00102/2007) of the Hungarian National Office for Research and Technology
 - Etocom project (TÁMOP-4.2.2-08/1/KMR-2008-0007)