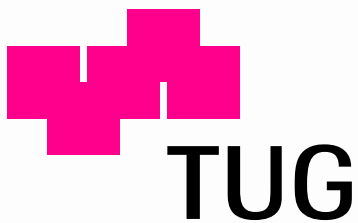# Speech Synthesis:
A Brief Motivation and Overview

Erhard Rank

TUG

Contents

- Speech synthesis?

- Examples

- Human speech production

- Human speech signals

- Synthesis concepts

- Summary

Speech synthesis?

Why is it of interest?

Speech communication

- is the most natural form of human communication

- is not bound to a display (can be used while driving a car/bike, working in adverse environment, etc.)

- many people today have an output terminal on them (mobile phone)

- ...

Speech synthesis?

What is it?

Generation of a speech signal by a machine

Fixed (small) number of utterances:

$\rightarrow$ Replay recorded speech (e.g., alarm system)

Arbitrary utterances ("unrestricted" vocabulary)

$\rightarrow$ Synthesize speech

– from rules using a production model

– using recordings of natural speech

Speech synthesis?

Flavors:

- Text-to-speech (TTS)
- Concept/content-to-speech (CTS)
- Dialogue system

Speech synthesis?

Flavors:

- Text-to-speech (TTS)

- Concept/content-to-speech (CTS)

- Dialogue system

Example: E-mail reading

S.g. Mag. Pospischil

Das MRT meeting findet am 16. schon um 8h im Sem.
217 statt, und nicht wie angekündigt um 10 ;(

MfG, Christian Fiala

Speech synthesis?

Flavors:

- Text-to-speech (TTS)
- Concept/content-to-speech (CTS)
- Dialogue system

Example: Weather forecast
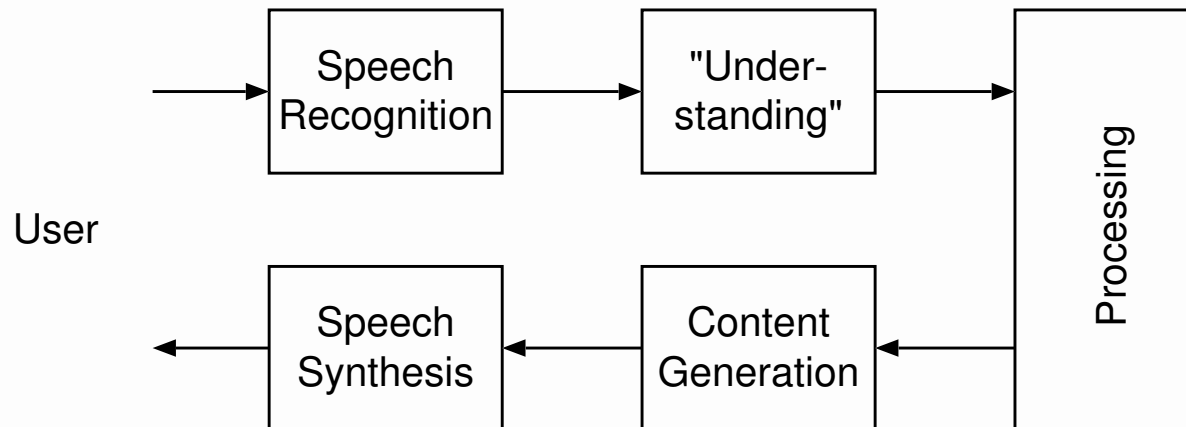
Wien: 17 Grad
Graz: 19 Grad
Linz: 13 Grad
. . .

$\Rightarrow$

"Das Wetter in Österreich beruhigt sich wieder. Am Nachmittag können allerdings nördlich des Alpenhauptkamms lokale Regenschauer. . . "

Speech synthesis?

Flavors:

- Text-to-speech (TTS)
- Concept/content-to-speech (CTS)
- Dialogue system

Speech synthesis?

Quality criteria:

- Intelligibility
- Naturalness

Examples (TTS)

- **AT&T nextgen**

  (1) Willkommen beim Sprachsynthesesystem Natural Voices von AT&T.

  (2) Meine Oma fährt im Hühnerstall Motorrad.

  (3) Der Zug von Grammatneusiedl nach Sopron fährt um viertel vor vier.

- **SVOX**

  (1) Meine Oma fährt im Hühnerstall Motorrad.

  (2) Wann geht der nächste Zug nach Venezia Mestre?

Examples (TTS)

- MBROLA, German

  (1) Die Börse: schwacher Wochenstart am deutschen Aktienmarkt.

  (2) Hallo, hört mich jemand? Schade, keiner da! Na, dann geh' ich halt nach Hause.

- Various (emotional synthesis)

  (1a) (1b) (1c) (1d) (1e)

  (2a) (2b) (2c) (2d) (2e)

  (3a) (3b) (3c) (3d)

Human speech production

Acoustic source

- oscillations of the vocal folds: voiced sounds (vowels or voiced consonants)

- noise due to turbulent air flow: unvoiced sounds

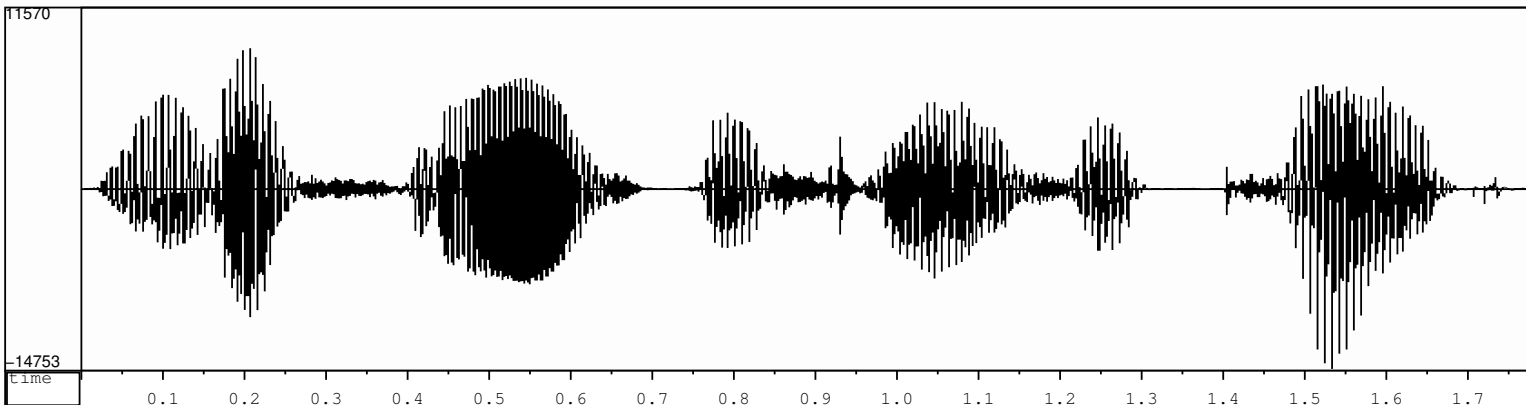- noise bursts due to sudden realease of closure: plosive sounds

Different positions of articulators (tongue, teeth, lips,. . .)

- different spectral shaping of speech sounds

Human speech signals
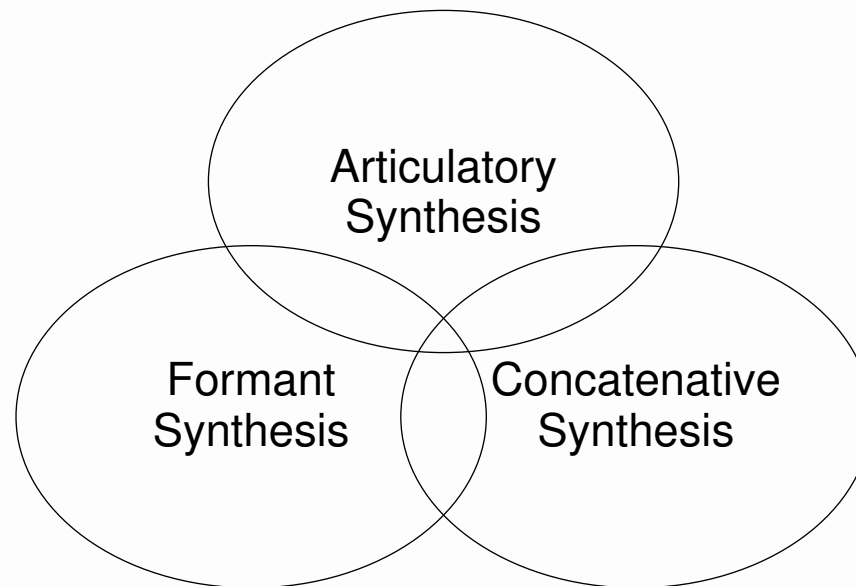

Live analysis...

Human speech signals

- Non-stationary

- Voiced (almost periodic) – unvoiced (noise-like)

- Realization of phonemes context dependent

- Prosody used for stressing, accentuation, contradiction,. . .

- Situation, dialogue partner, emotion, etc.

Synthesis concepts

Categories:

Synthesis concepts
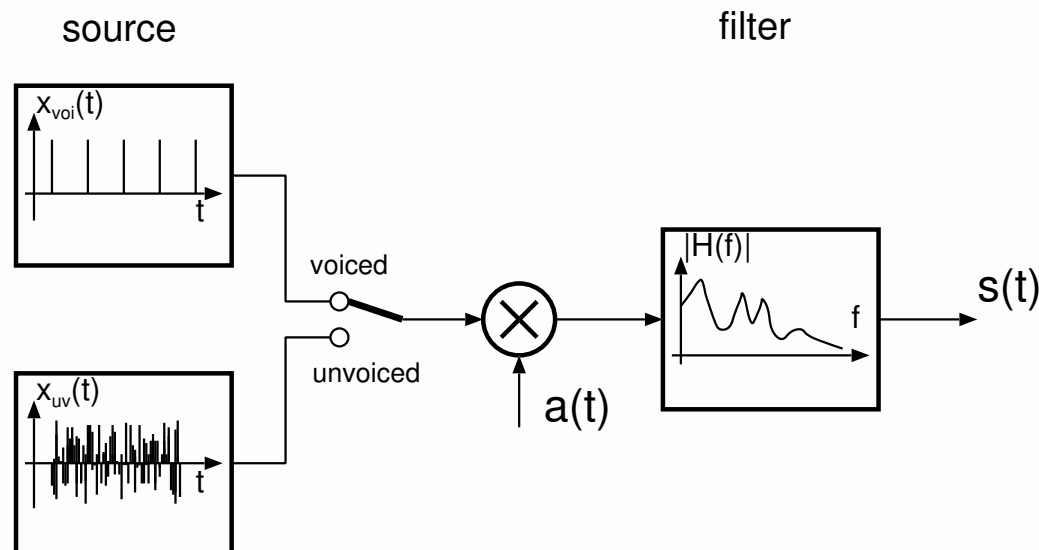
Physical Modeling (Artic. Synth.):

- Signal generation by a physical model of human speech production system
    - Vocal folds (oscillatory source)
    - Vocal tract (filtering, turbulent noise sources)
- Coupling between source and filter!
- Model parameters rely on (inaccurate) measurements
- Control of parameter trajectories difficult

Synthesis concepts

Formant synthesis:

- Establish formants by resonance filters
- Excited by a periodic (voiced) or noise (unvoiced) source signal

Synthesis concepts

Articulatory and formant synthesis:

Rule-based synthesis

However:

- Often the rules have to be derived from natural recorded signals or measurements (X-ray)

Alternative:

Database-driven synthesis

Synthesis concepts

Database-driven (concatenative) synthesis:

Generate speech by concatenation
of recorded elements

- Dedicated inventory
    - Phonemes, di-phones, demi-syllables,...
    - Prosodic manipulations
- Large inventory: Unit selection
    - Selection optimal signal units
    - No/few signal manipulations

Summary

- Speech synthesis: Why, what, how?
- Quality: Intelligibility vs. naturalness
- Audio examples
- Human speech production/signals
- Synthesis concepts