Emotional Speech Synthesis

State of the art 2009

Felix Burkhardt

outline

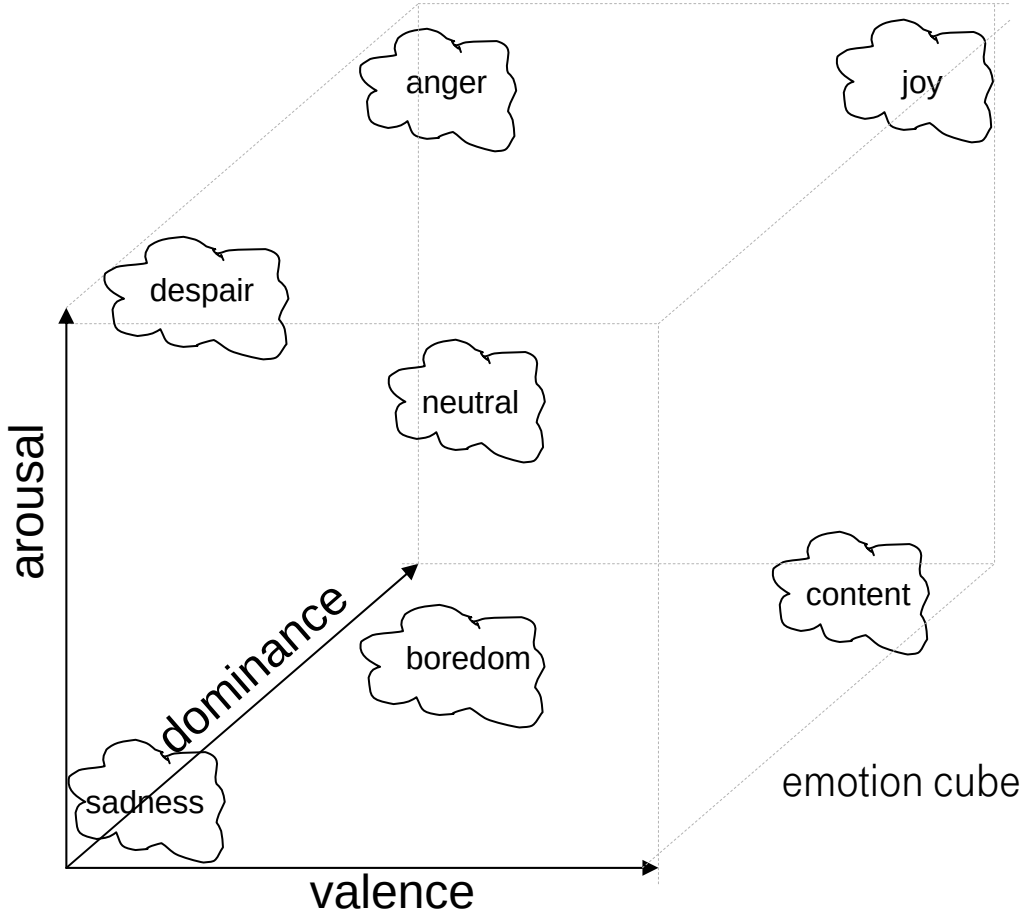
- how to model and why simulate emotions?
- emotions in speech
- introduction to speech synthesis approaches
- examples, examples, examples
- conclusion and outlook

contents

- how to model and why simulate emotions?
- emotions in speech
- overview on speech synthesis
- examples, examples, examples
- conclusion, outlook

emotion models

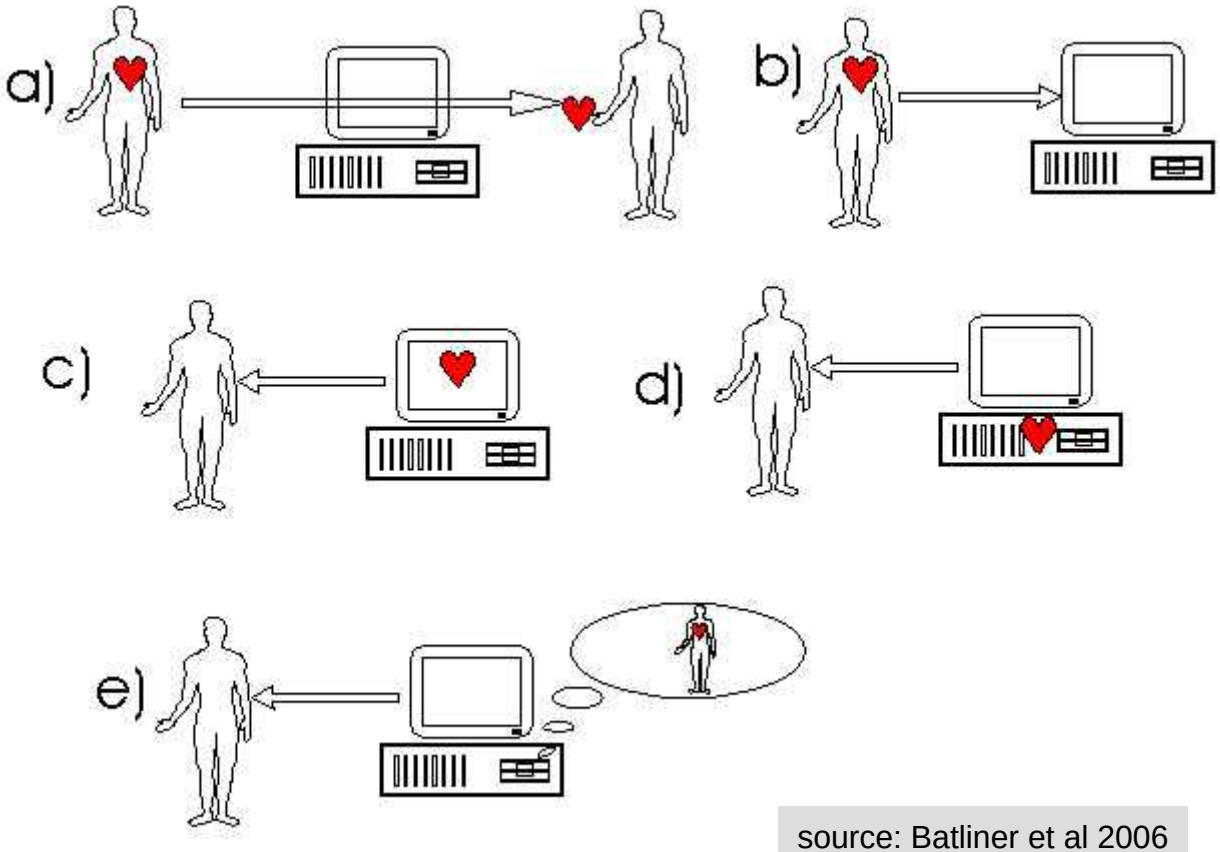
- …everyone except a psychologist knows what an emotion is (Young 1973)
- categories, e.g. anger, joy, …
- dimensions, e.g. activation, dominance, valence
- appraisals, e.g. novelty, intrinsic pleasantness, relevance, coping potential,



source: Burkhardt 2001

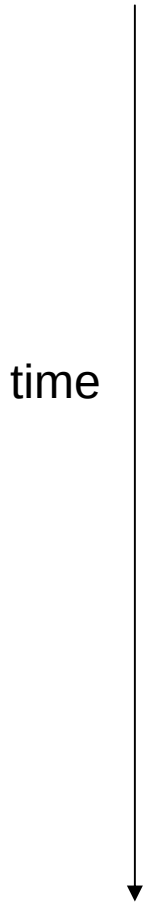
why model emotional behaviour?

- aspects of emotion modeling in human-machine interaction:



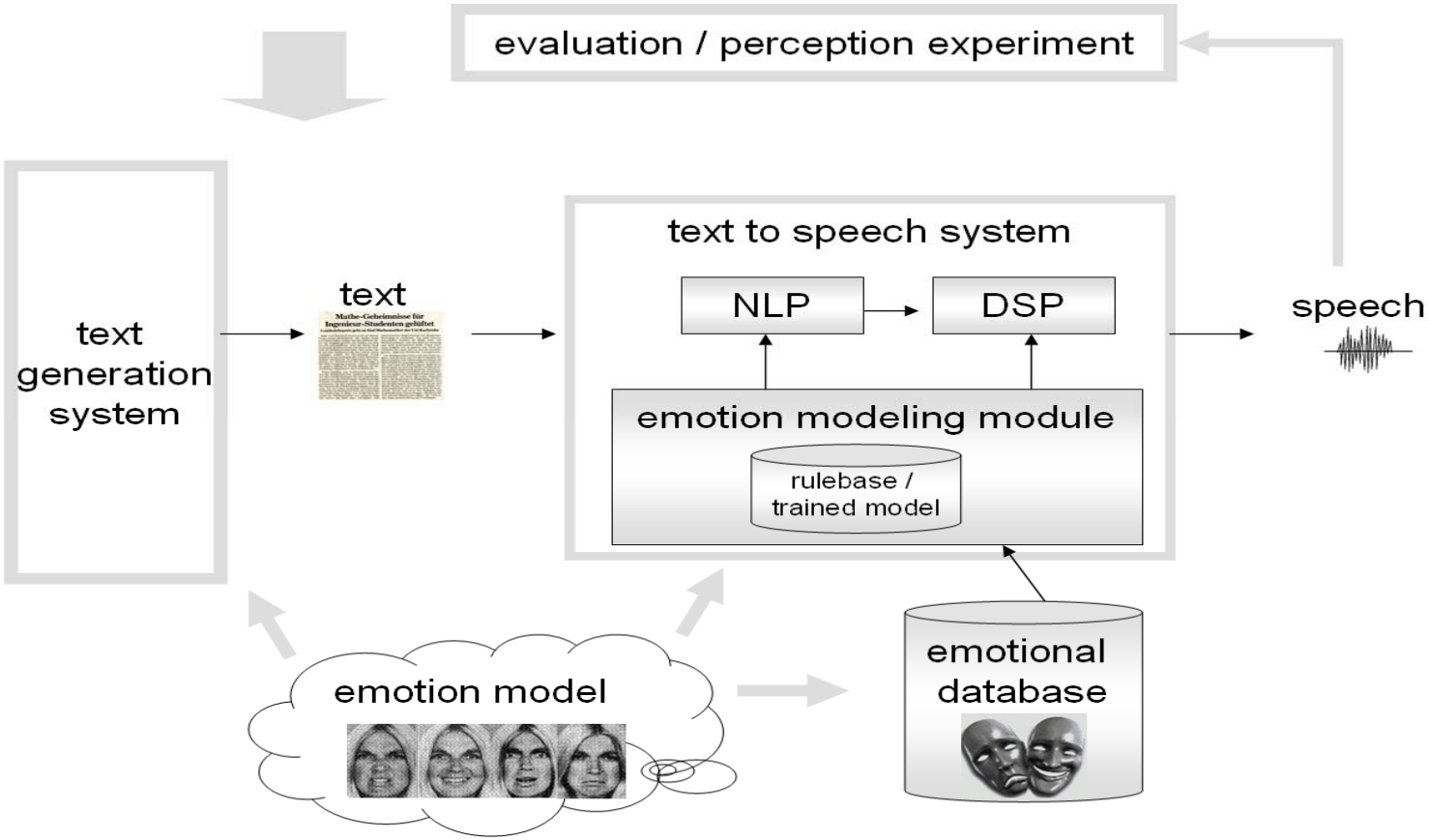
source: Batliner et al 2006

applications of emotional tts



- fun, e.g. emotional greetings
- prosthesis
- emotional chat avatars
- gaming, believable characters
- adapted dialog design
- adapted persona design
- target-group specific advertising
- …
- believable agents
- …
- artificial humans

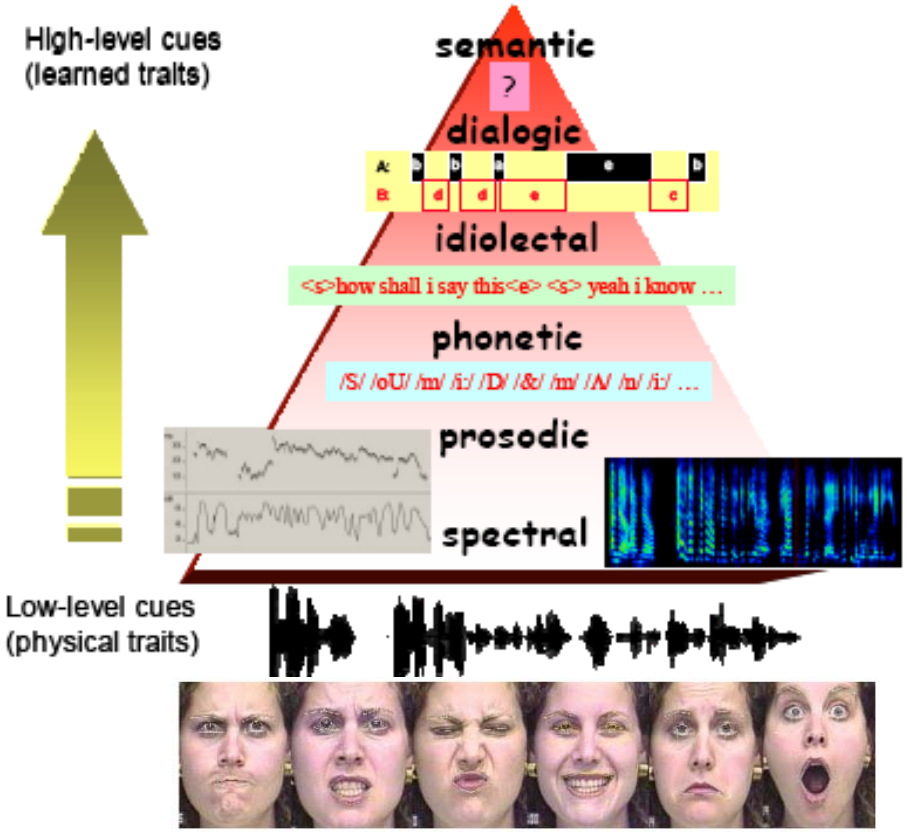
aspects of emotional tts



contents

- why simulate emotions?
- emotions in speech
- overview on speech synthesis
- examples, examples, examples
- conclusion, outlook

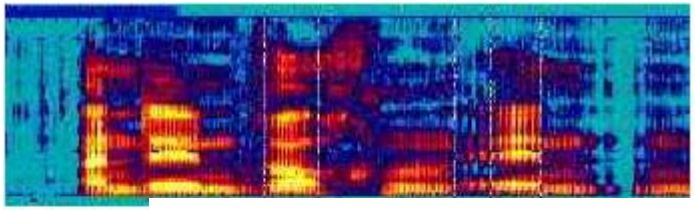
speech features



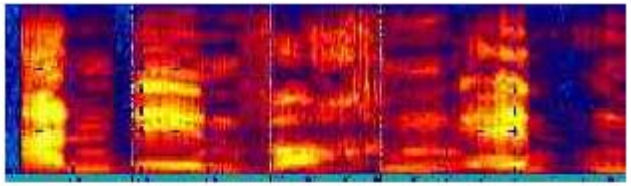
descriptive layers of speech

source: Reynolds et al 2003

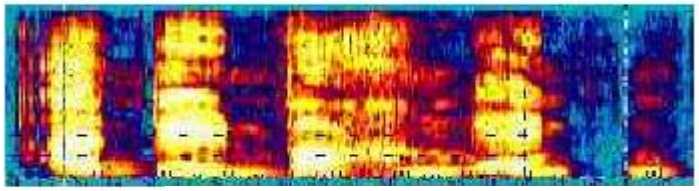
emotion in speech



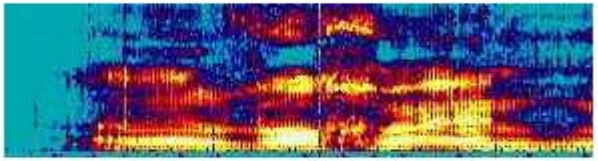
neutral



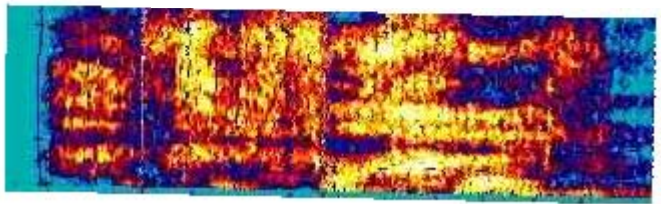
angry



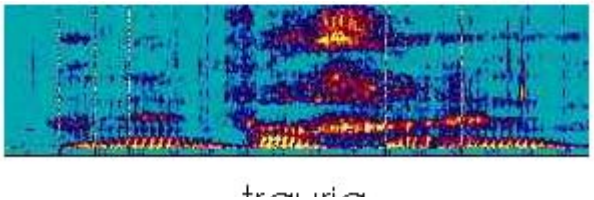
happy



bored



frightened



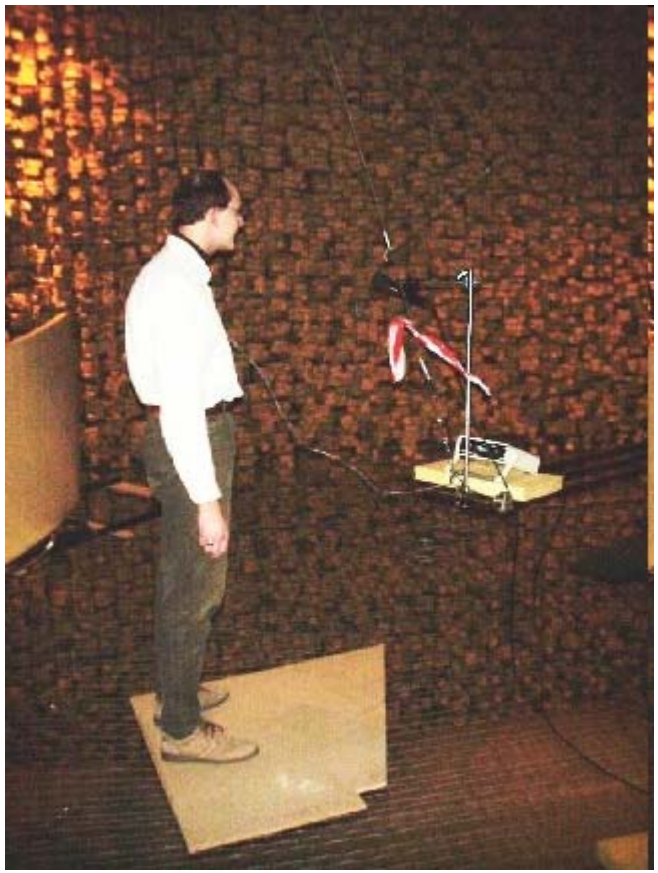
sad

spectrograms from emotional acted speech

source: TUB emotional database

emotional data?

- actors vs. reality
- Berlin EmoDB: 10 actors x 7 emotions x 10 sentences
- alternatives
  - induced data, e.g. Aibo
  - television, radio data



EmoDB: Burkhardt et al 2005

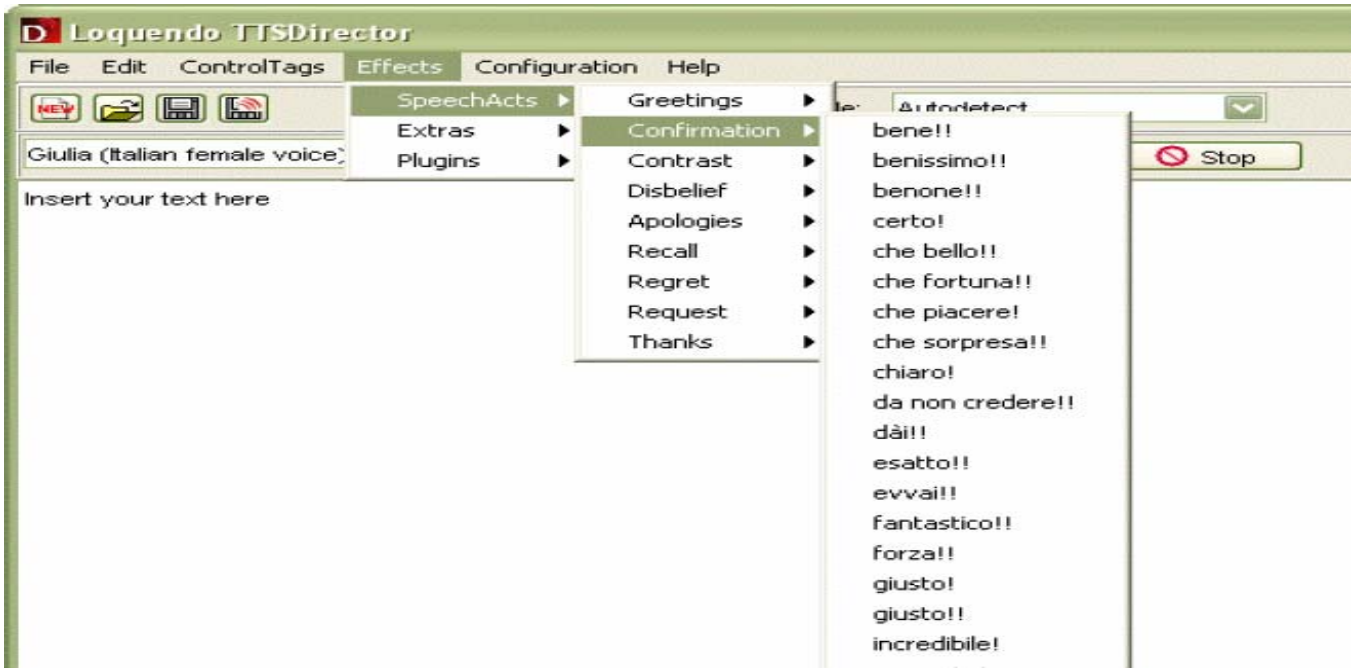
how to describe emotion?

- EmotionML, incubator group at W3C
- Example, embedded in SSML:

```
<speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  <voice gender="female">
    <prosody contour="(0%,+20Hz)(10%,+30%)(40%,+10Hz)">
       Hi, am sad know but start getting angry...
    </prosody>
  </voice>
  <emotion>
   <category name="sadness„ set="basic" intensity="0.6"/>
   <timing start="10%" end="50%"/>
  </emotion>
  <emotion>
   <category name="anger" set="basic" intensity="0.4"/>
   <timing start="50%" end="100%"/>
  </emotion>
</speak>
```

http://www.w3.org/2005/Incubator/emotion/

loquendo tts director

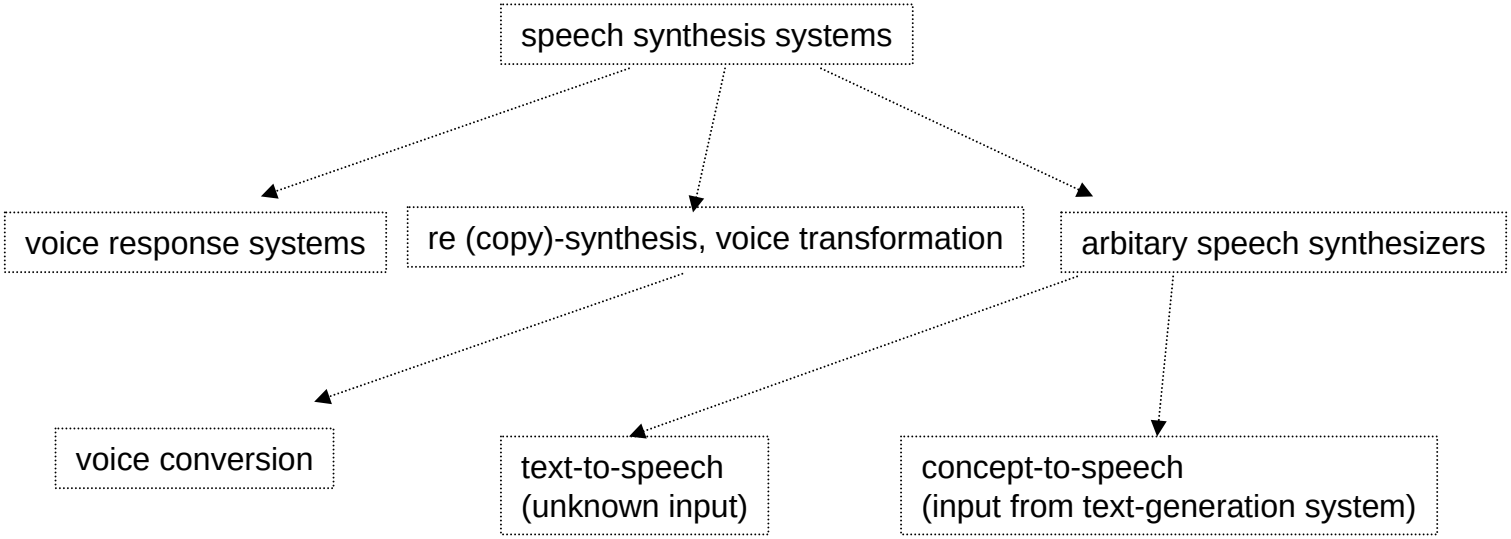


source: Loquendo

contents

- why simulate emotions?
- emotions in speech
- introduction to speech synthesis approaches
- examples, examples, examples
- conclusion, outlook

speech synthesis taxonomy



tts process chain

NLP natural language processing

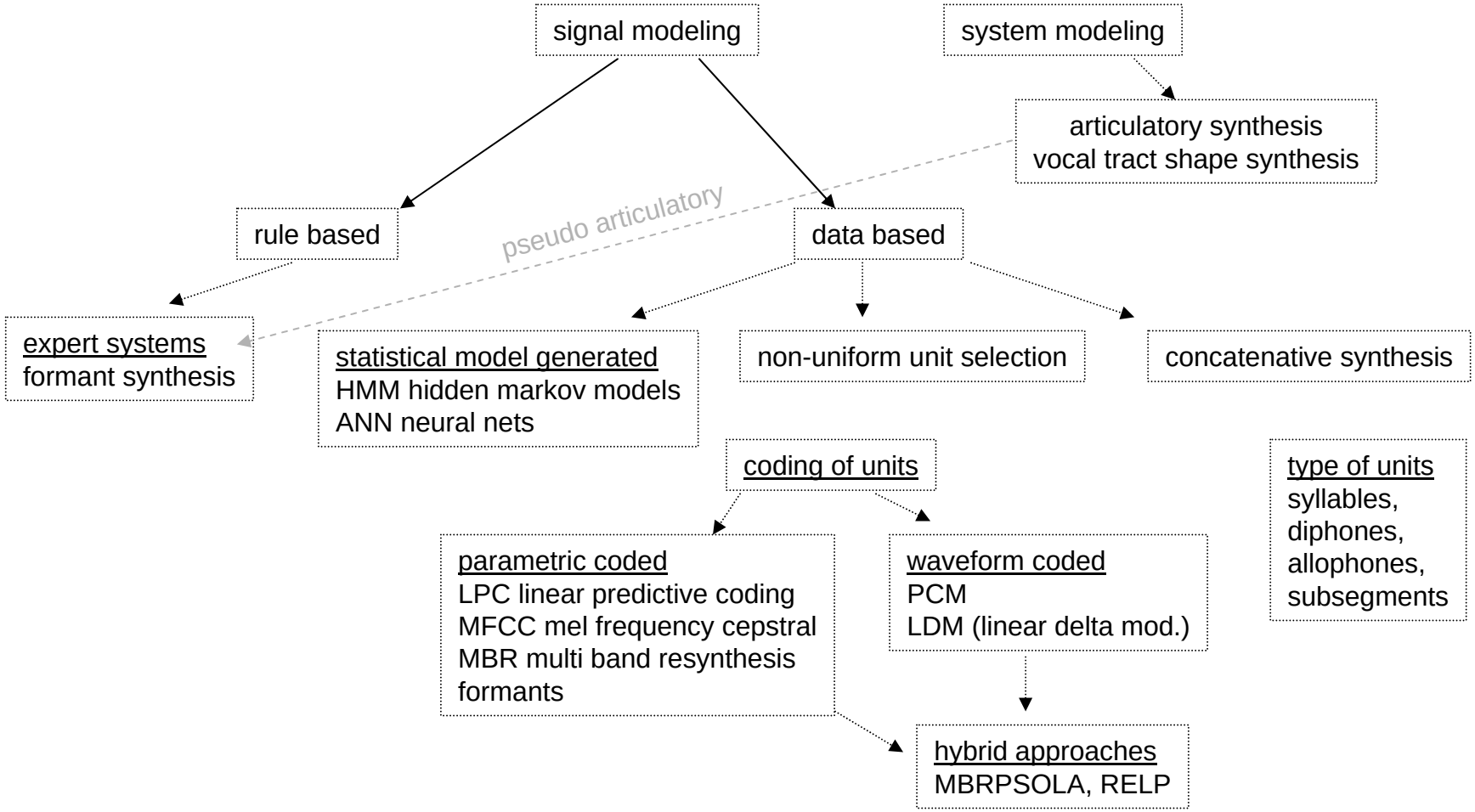
phonetic transcription
prosody track

DSP digital speech processing

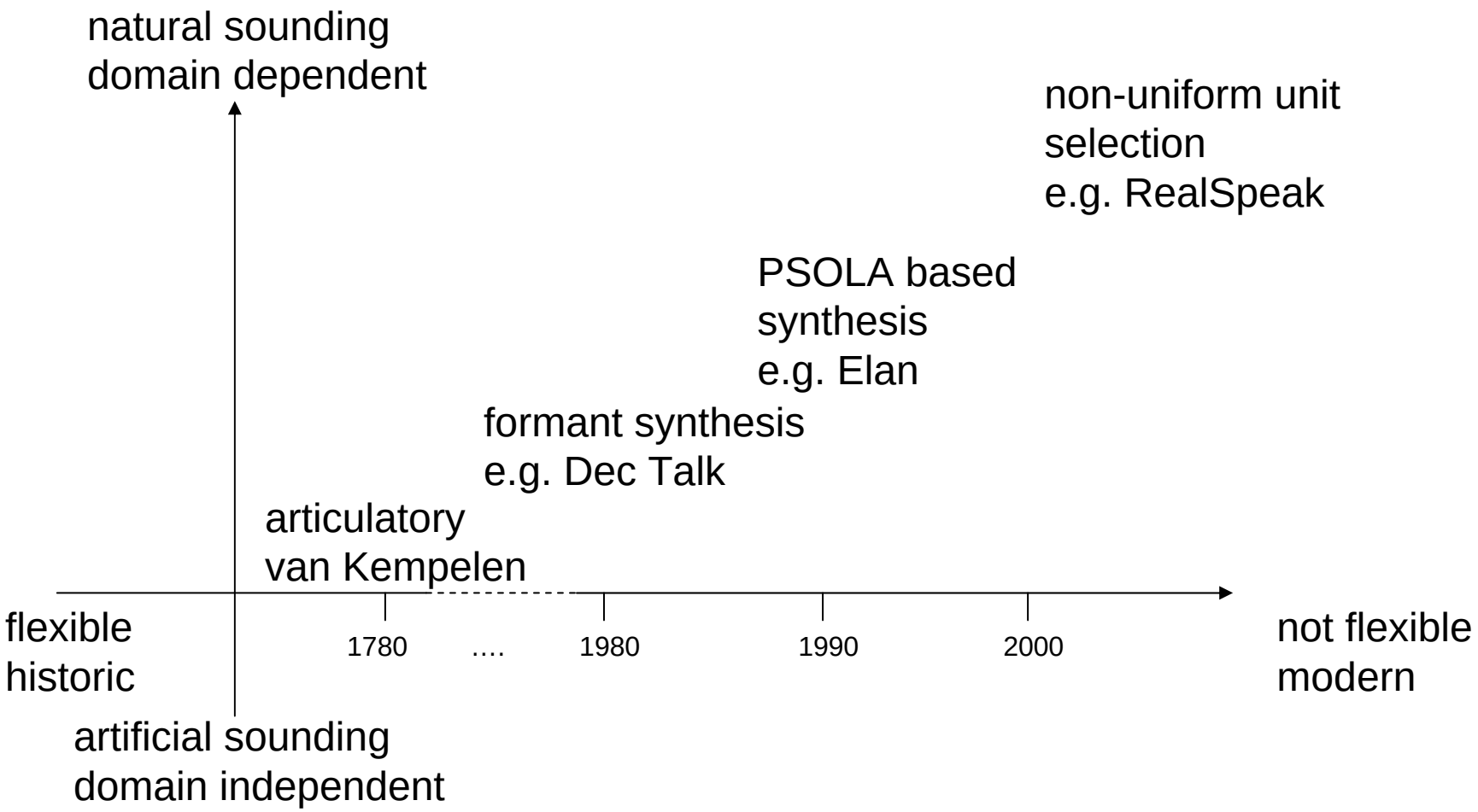
preprocessing
morpho-syntactic analysis
transpcription
prosody modeling

unit concatenation / search
prosody fitting
edge smoothing

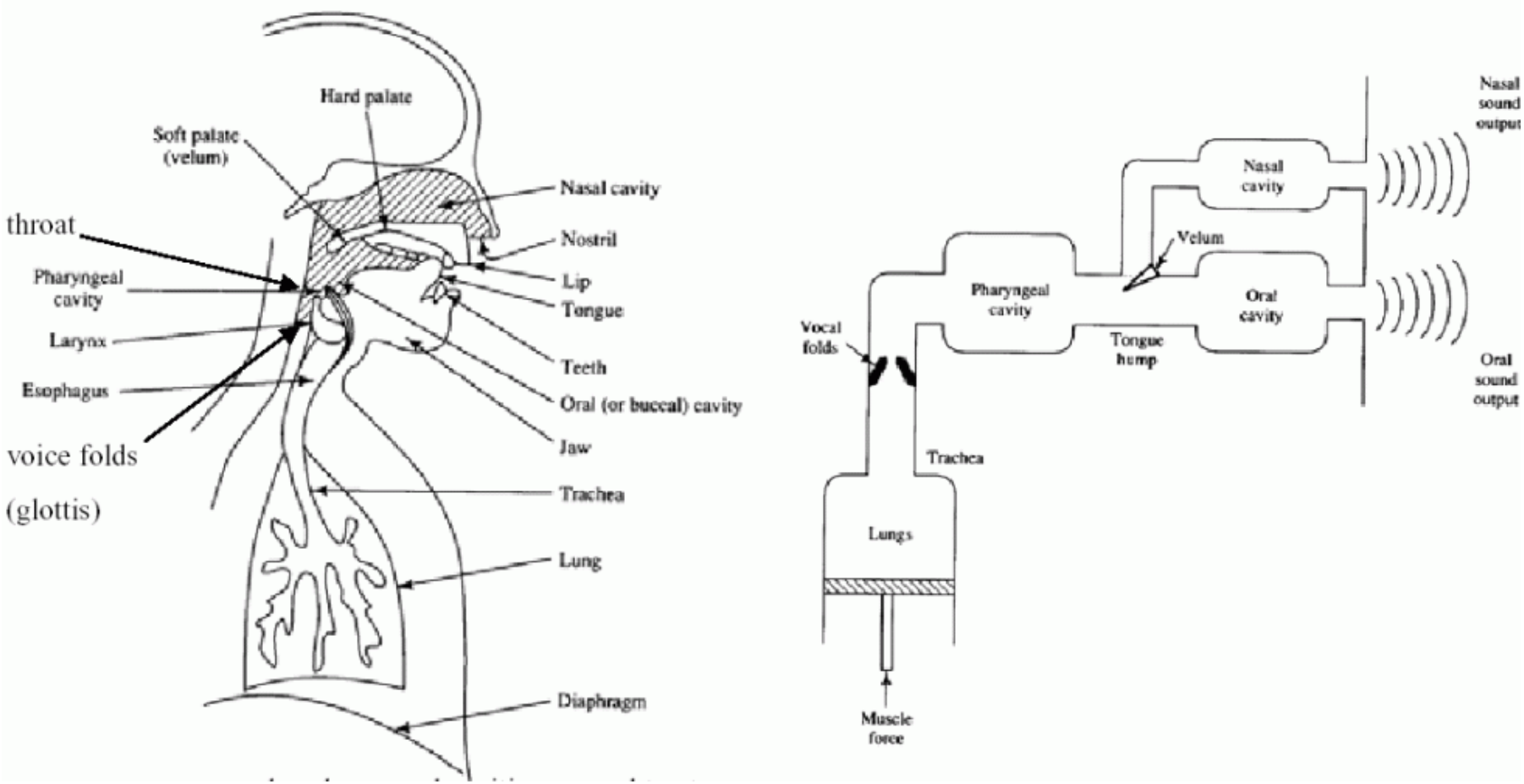
synthesis approaches



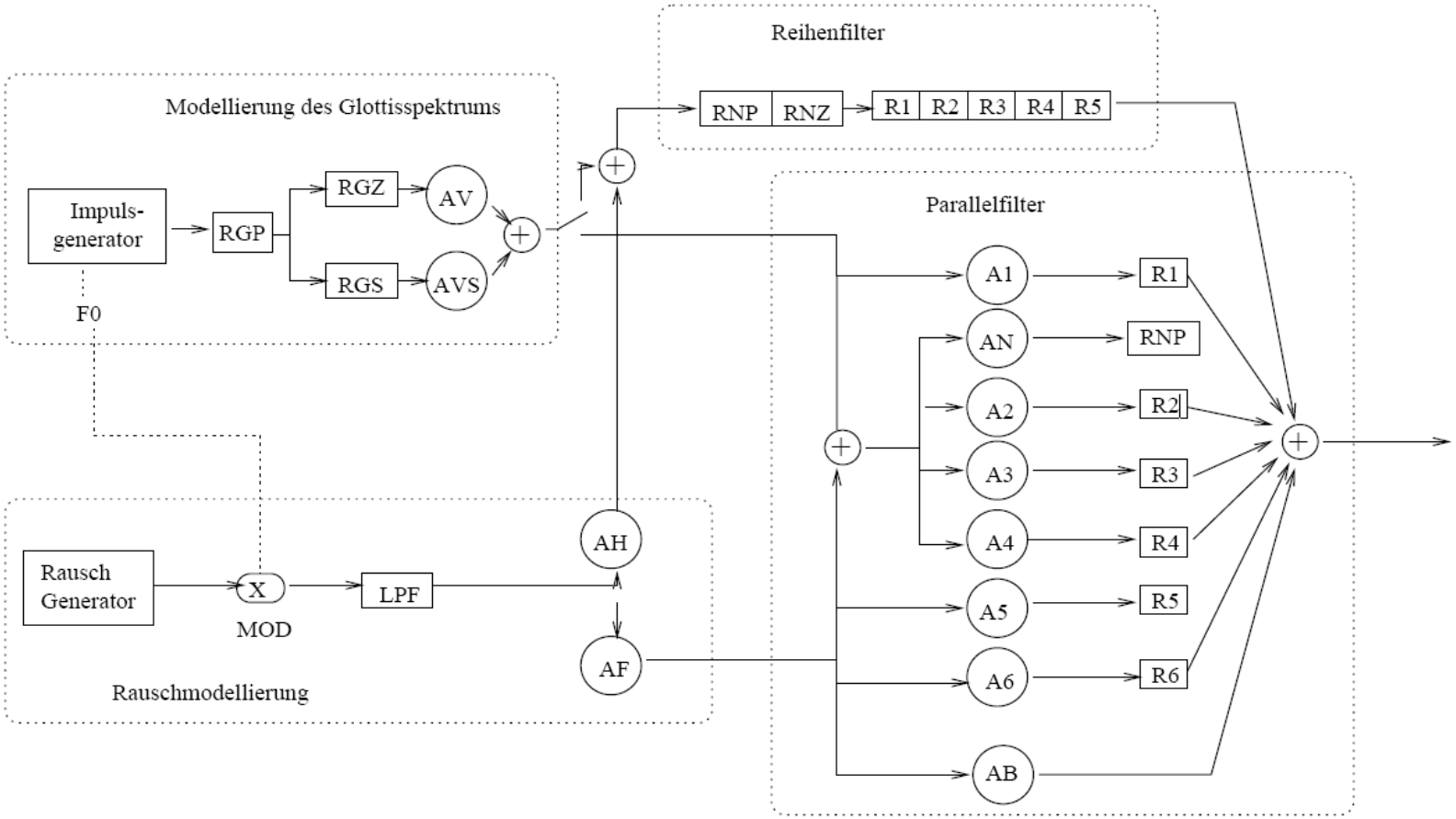
historic development



system modeling



source filter model



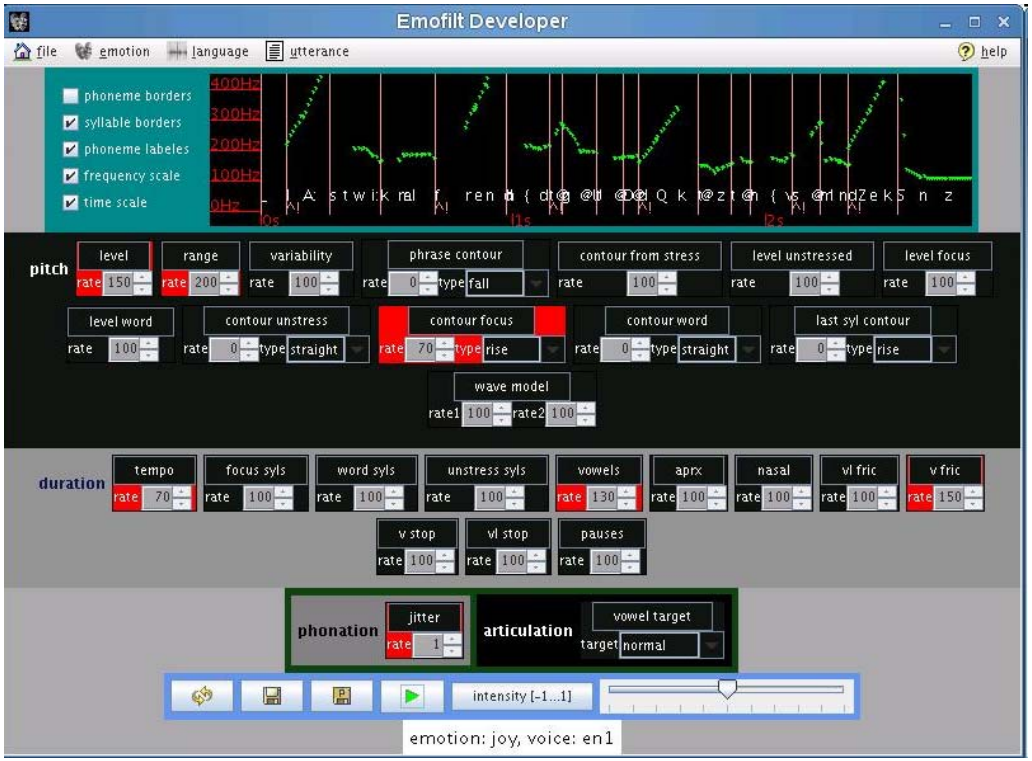
source: Klatt80 formant synthesizer (Klatt 1980)

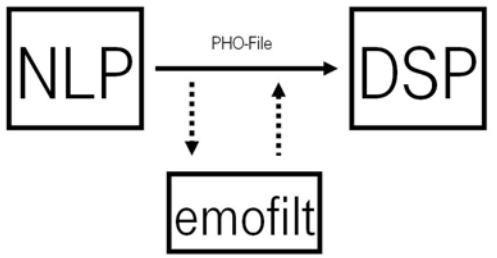
contents

- why simulate emotions?
- emotions in speech
- overview on speech synthesis
- examples, examples, examples
- conclusion, outlook

examples: emofilt

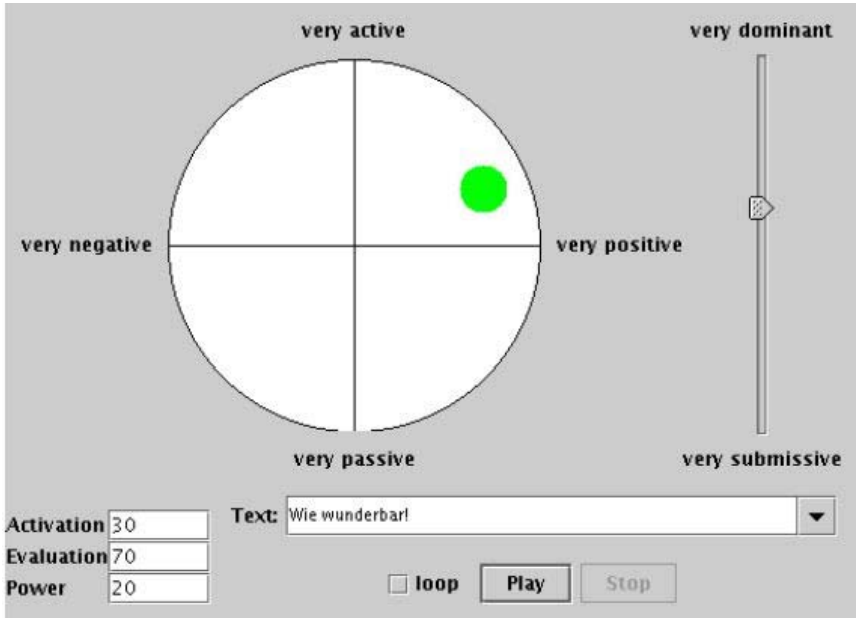
- open source Java program based on MBROLA synthesis engine.
- NOT a complete text-to-speech system
- prosody filter between natural language and digital speech signal processing modules
- as multilingual as MBROLA which currently supports 35 languages.





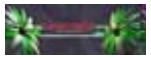
examples: emoSpeak

- emoSpeak is integrated into the MARY text-to-speech framework by DFKI.
- Marc Schröder investigated in his ph.d. thesis, how to assign rule-based modification of speech to emotional dimensions.
- the system can be freely dowloaded



source: Schröder 2004

examples voice conversion

Murtaza Bulut et al, USC 	PSOLA - LPC conversion	neutral angry
Greg Beller, IRCAM 	Phase vocoder	neutral sad

examples voice transformation

Olivier Rosec FranceTelecom 2009	Mixed LF + harmonic model	woman as boy as man man breathy whispery tense
Shiva Sundara USC 2007	Laughter synthesis by LPC synthesis and mass-spring model	

examples formant synthesis

AffectEditor  J. Cahn, MIT 1998	DEC Talk prosody rules	sad angry
EmoSyn  Burkhardt, 2000	prosody rules + phonation model	neutral sad angry crying content

examples diphone synthesis

MARY M. Schröder, DFKI	prosody rules for dimensions three inventories for soft, normal and tense speech	joy	angry
EmoFilt Burkhardt, 1999	prosody rules	neutral	joy

examples statistical based

Tokyo Institute,
Kobayashi Lab

HMM models spectral and prosodic features

neutral joy

examples unit selection

Cepstral — fun personality voices — Damian Shouty


IBM — CTTS with expressive units — product research


Loquendo — extralinguistic units — Katrin

examples non human

Oudeyer: Sony pet robot 	concatenative	happy sad
MIT Kismet robot 	formant synthesis	anger fear

examples singing

vocal tract lab Peter Birkholz		2007 articulatory	donna nobis
pavarobotti Ingo Titze		1993 Articulatory	aria
Bell Labs Gerstman & Mathews,		1961 articulatory, first song ever	bicycle

more examples ...

http://emosamples.syntheticspeech.de

Expressive Synthetic Speech


(pictures taken from P. Ekman)

last update: May 4th 2009

This is a collection of examples of synthetic affective speech conveying an emotion or natural expression and maintained by Felix Burkhardt. Some of these samples are direct copies from natural data, others are generated by expert-rules or derived from data-bases. The emotional labels "anger", "fear", "joy" and "sad" are my (short) designators for "the big four" basic emotions, not neccessarily the authors' ones.
Examples of German actors simulating emotional arousal can be found here.
Examples of German text-to-speech synthesizers can be found here.
Please, feel encouraged to let me know about own or missing attempts to simulate emotional speech! (felixbur@gmx.de)

contents

- comparing simulation of anger, fear, joy and sadness
- other simulations
- related examples
- further links
- projects concerning emotional speech
- changelog

All Audiofiles are Mp3-format (64 or 32 kB/s)

author	visual	affil.	year (approx)	description	neutral	anger	joy	sad	fear
N. Audibert, V. Aubergé, A. Rilliard	Institut de la Communication Parlée	ICP	2006	Copy prosody and intensity from satisfied and sad speech to neutral speech using PSOLA technique. See the article "The Prosodic Dimensions of Emotion in Speech: the Relative Weights of Parameters", Interspeech 2005 (Lisbon), for details	▶	-	▶	▶	-
Stephan Baldes	DFKI	DFKI	1999	Rule based emotion simulation with Entropic's formant TTS engine TrueTalk, based on Cahn's affect editor approach	-	▶	-	▶	▶

contents

- why simulate emotions?
- emotions in speech
- overview on speech synthesis
- examples, examples, examples
- conclusion, outlook

conclusion

- emotions are part of natural speech
- simulation possible by either
  - modeling the process
  - including emotional data
- still text to speech fights with intelligible, neutral speech
- first steps: speaking styles, extralinguistics
- first apps: fun, gaming

outlook

- discrepancy between
  - natural but unflexible vs.
  - artificial sounding but flexible
- solutions short - middle term:
  - very large databases
  - hybrid parametric – non-uniform unit selection
  - voice transformation techniques
  - high quality source filter model based synthesis
- solutions on the long run
  - physical modeling

references

- http://emosamples.syntheticspeech.de/
- F. Burkhardt, „Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren“, Shaker Verlag 2001
- Burkhardt & Sendlmeier, “Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis”, Proc. ISCA Workshop on Speech and Emotion, 2000,
- A. Batliner, F. Burkhardt, M. van Ballegooy, E. Nöth: A Taxonomy of Applications that Utilize Emotional Awareness, Proc. IS-LTC 2006
- Schröder, M. (2001). “Emotional Speech Synthesis - A Review”. Proc. Eurospeech 2001
- Dutoit, T., “An Introduction to Text-to-Speech Synthesis”, Springer 1997
- Murray & Arnott, “Synthesizing emotions in speech: Is it time to get excited”, 1996
- Tokuda et al, “An HMM-based speech synthesis system applied to English”, IEEE Workshop on Speech Synthesis 2002
- Agiomyrgiannakis, Olivier Rosec: "ARX-LF-based source-filter methods for voice modification and transformation", Proc. ICASSP 09
- Murtaza Bulut et al ,"Investigating the role of phoneme-level modifications in emotional speech resynthesis", Proc Interspeech 2005
- Ellen Eide et al, “A Corpus-Based Approach to <AHEM/> Expressive Speech Synthesis”
- Oudeyer P-Y. "The Synthesis of Cartoon Emotional Speech", Proc.Int. Conference on Prosody
- Sundaram, s. and Narayanan, S., "Automatic acoustic synthesis of human-like laughter", Jasa No1, p. 527-535, 2007.
- Reynolds, D.; Campbell, J.; Campbell, B.; Dunn, B.; Gleason, T.; Jones, D.; Quatieri, T.; Quillen, C.; Sturim, D.; Torres-Carrasquillo, P. (2003). Beyond Cepstra: Exploiting High-Level Information in Speaker Recognition, Workshop on Multimodal User Authentication.
- Schröder, M. (2004). Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis
- http://www.w3.org/2005/Incubator/emotion/
- http://emotion-research.net/