universitätbonn

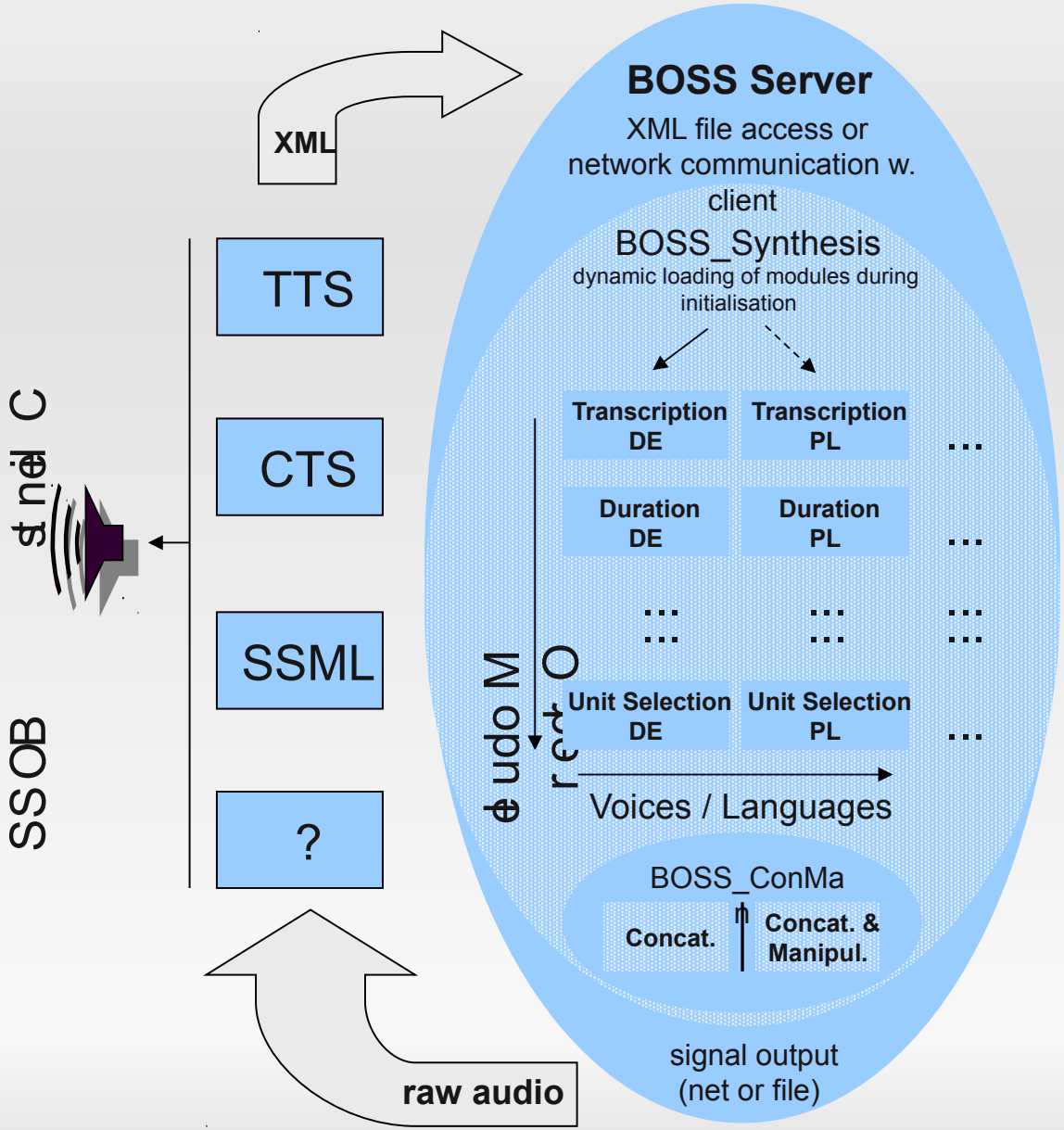
Universität Bielefeld

BOSS, Rap music, and a Dynamic oscillation model in human-machine communication

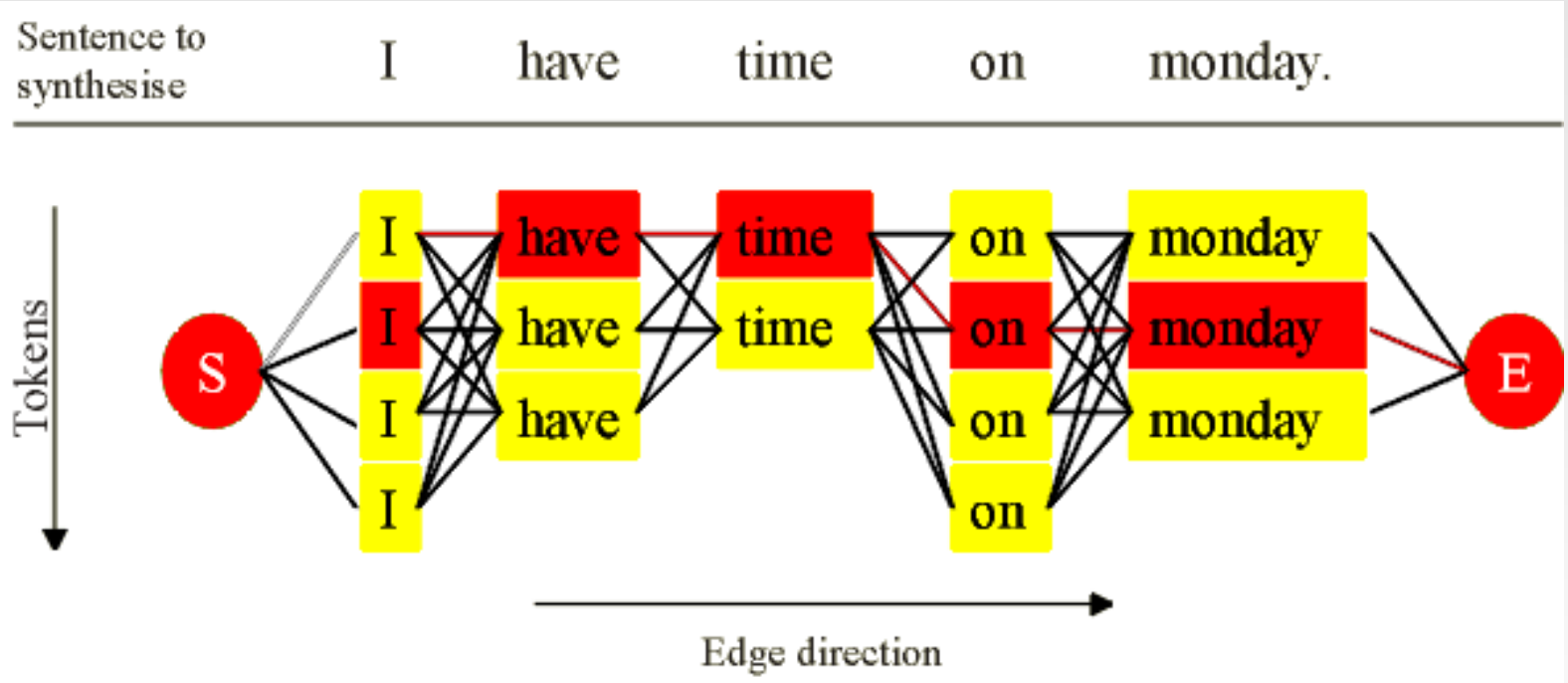
BOSS

- A Non-uniform unit-selection speech synthesis system developed at the institute of communication research and phonetics in the University of Bonn
- Designed for research: very modular
- Flexible in adaptation to new languages and adding new modules
- Open source
- Used in a project of the German Telecome in an automatic phone answering system

General Architecture



Unit selection process

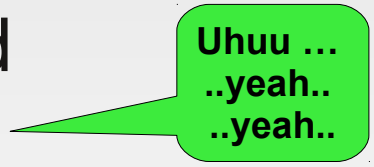


BOSS

- My work:
  - General development (adaptation to new systems, error search etc)
  - Module development:
    - F0 pattern classification algorithm based on temporal correlation
    - Prosody prediction based on classification and regression trees (CART) and hidden markov models (HMM)
    - Prominence prediction (rule based)
    - Adaptation for new synthesis units (PHOXSY) and diphone synthesis
    - Corpus integration
    - Development of a corpus integration toolkit (BOFU)
  - Representation in ECESS (European center of excellence in speech synthesis)

Rap music

- Master project
- Project goals:
  - Formal description of speech-music-timing relationship
  - Building a formal model: a dynamic perception-production mechanism
- Model evaluation: statistics and "audience" feedback modeling

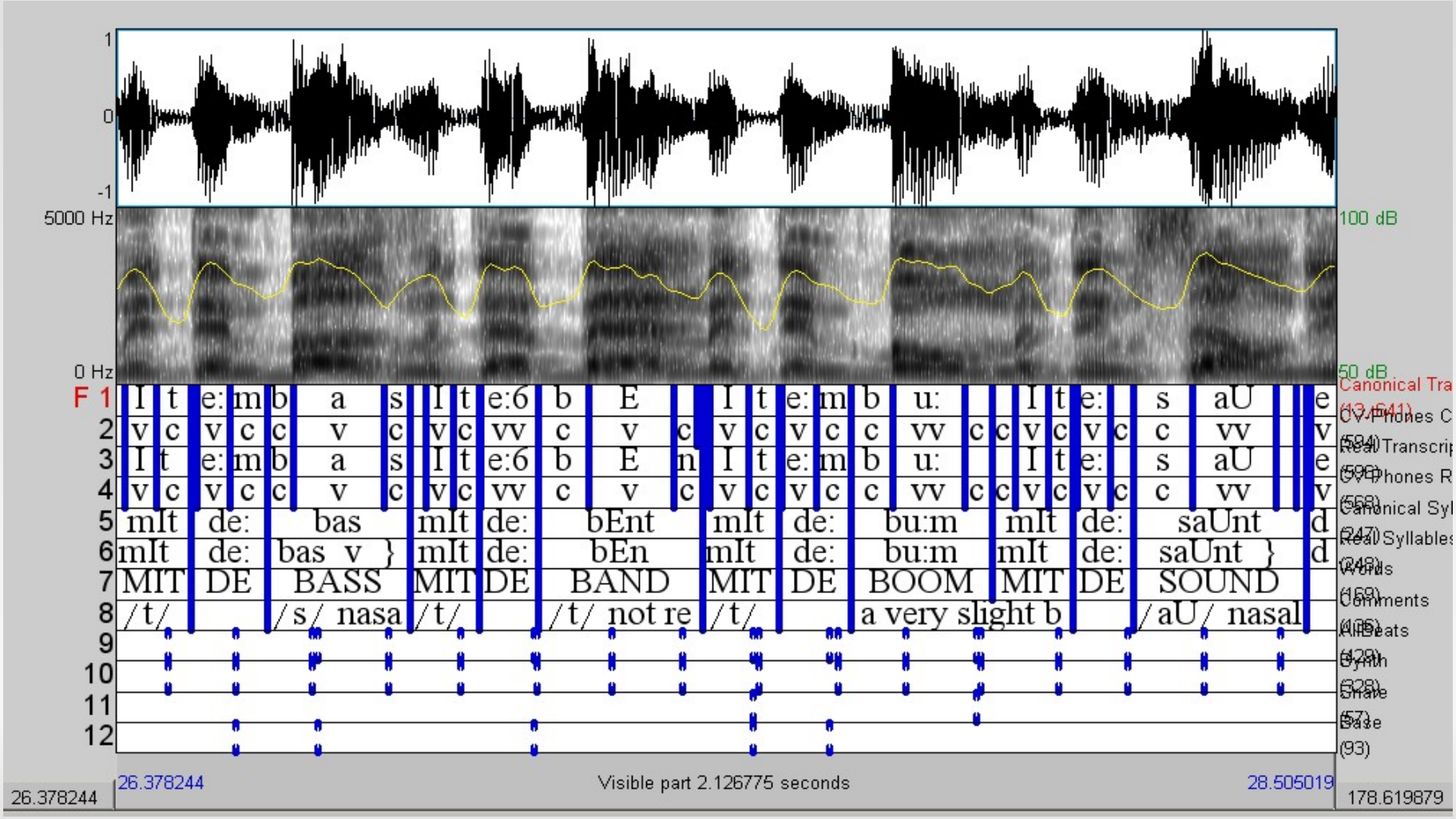


- Integrating the model into a rapping artificial agent

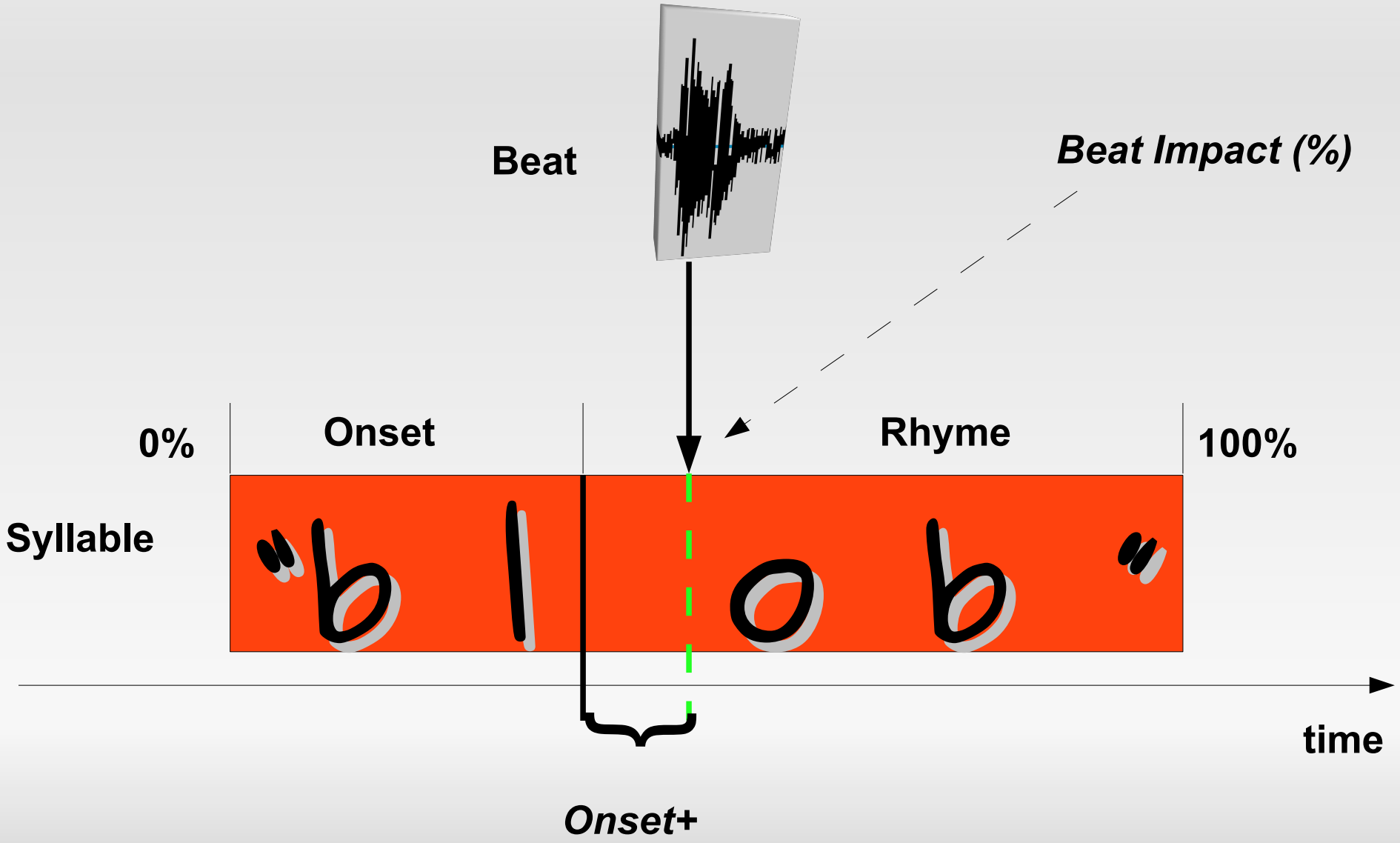
  → rap synthesizer

Material

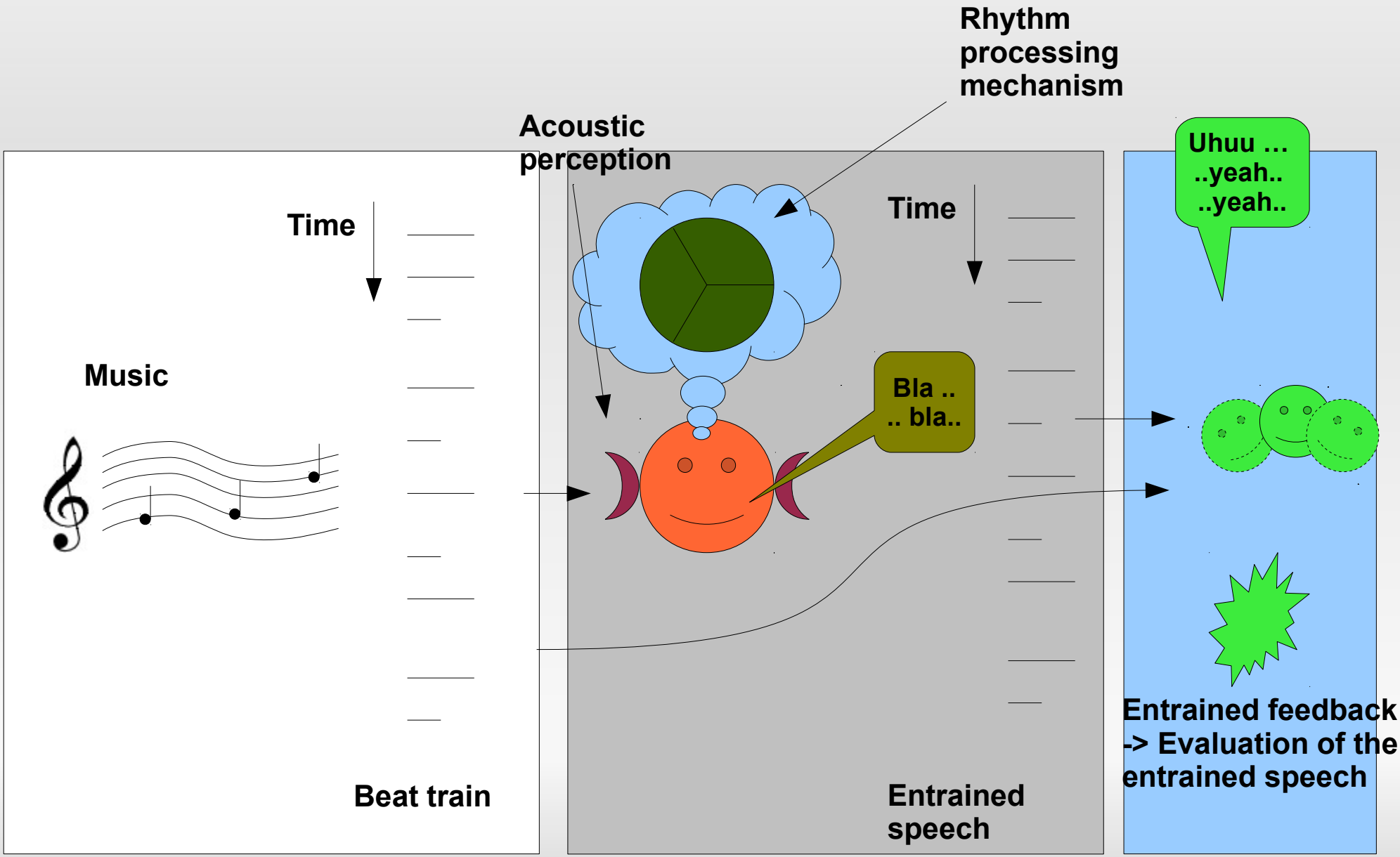
- 4 German rap songs performed by 3 professional artists



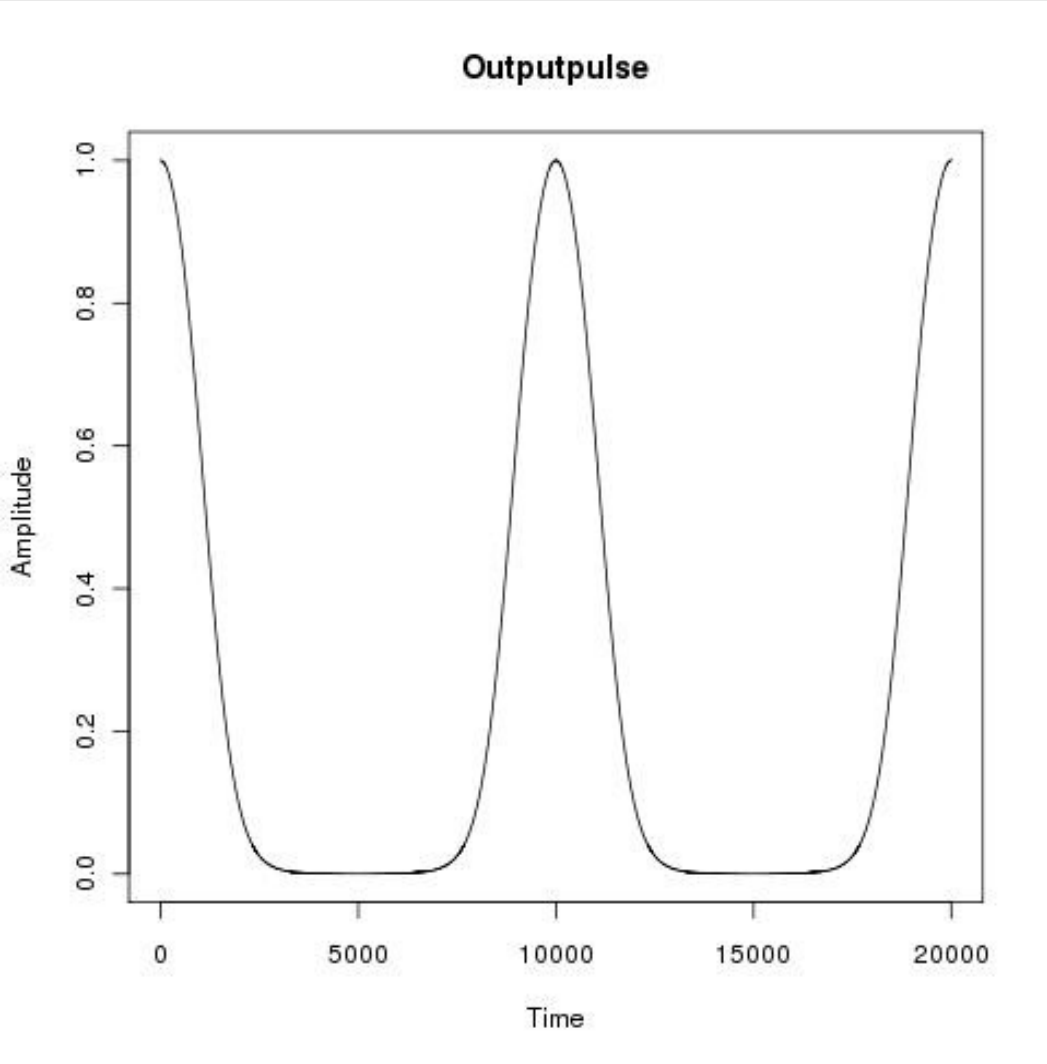
Basic terms: Beat Impacts and Onset+



Scenario


Rhythm processing mechanism
Acoustic perception
Time
Music
Beat train
Time
Bla .. .. bla..
Entrained speech
Uhuu … ..yeah.. ..yeah..
Entrained feedback -> Evaluation of the entrained speech

Global level: oscillator model output function

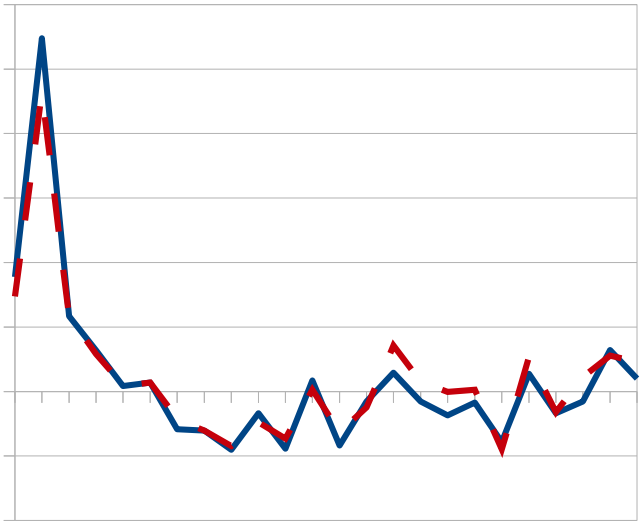


Large (1995)

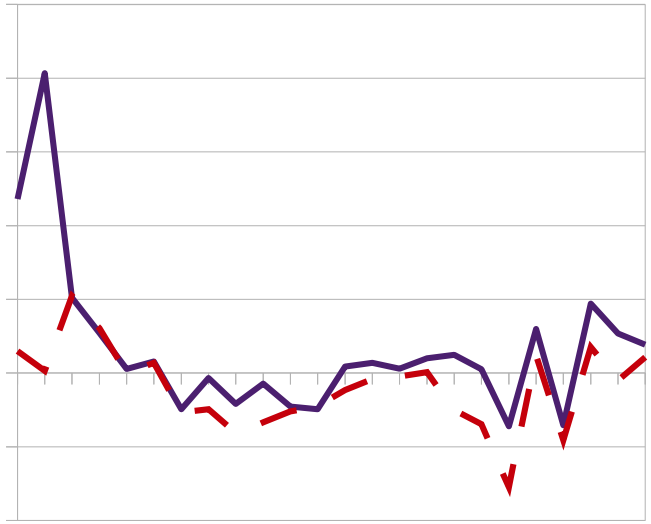
Statistic analysis results

- Beat Impact mean values remain stable at 0.5 (50%)
- Beat Impact standard deviations remain stable at 0.3 (30%)
- Correlations of P-Center deviations with Onset+ values between 0.9 and 0.96
- Other correlations appear to be individual for each song
- Bunch of correlations defined as a rhythmic song profile
- Idea: in case of optimal rhythmic representation by the oscillators song profiles of original rhythms and reproduced rhythms should be similar

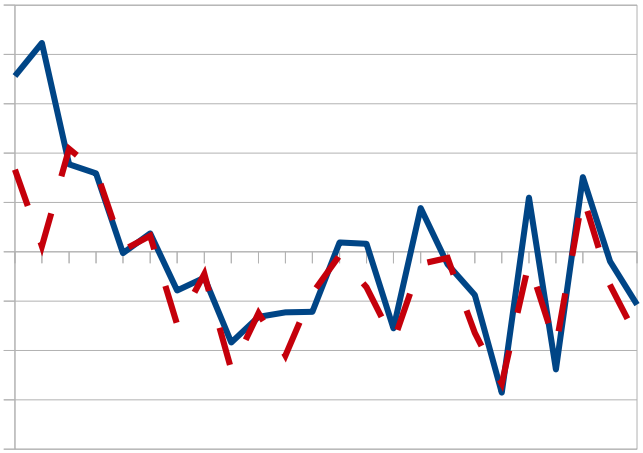
Global level: Song profile replication



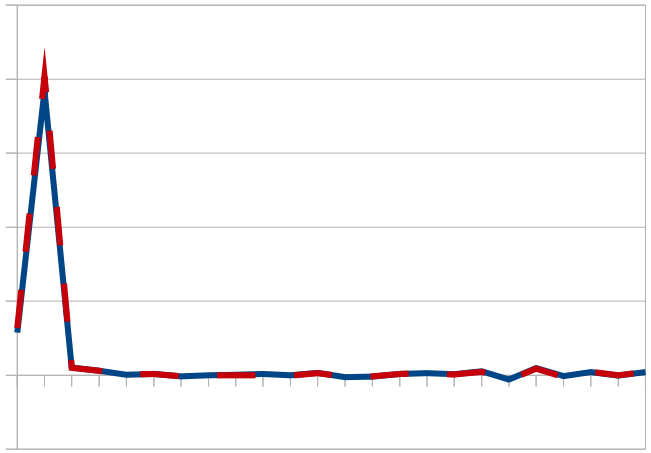
Base original
1/4 Oscillator response



Snare original
1/16 Oscillator response



Base original
1/8 Oscillator response

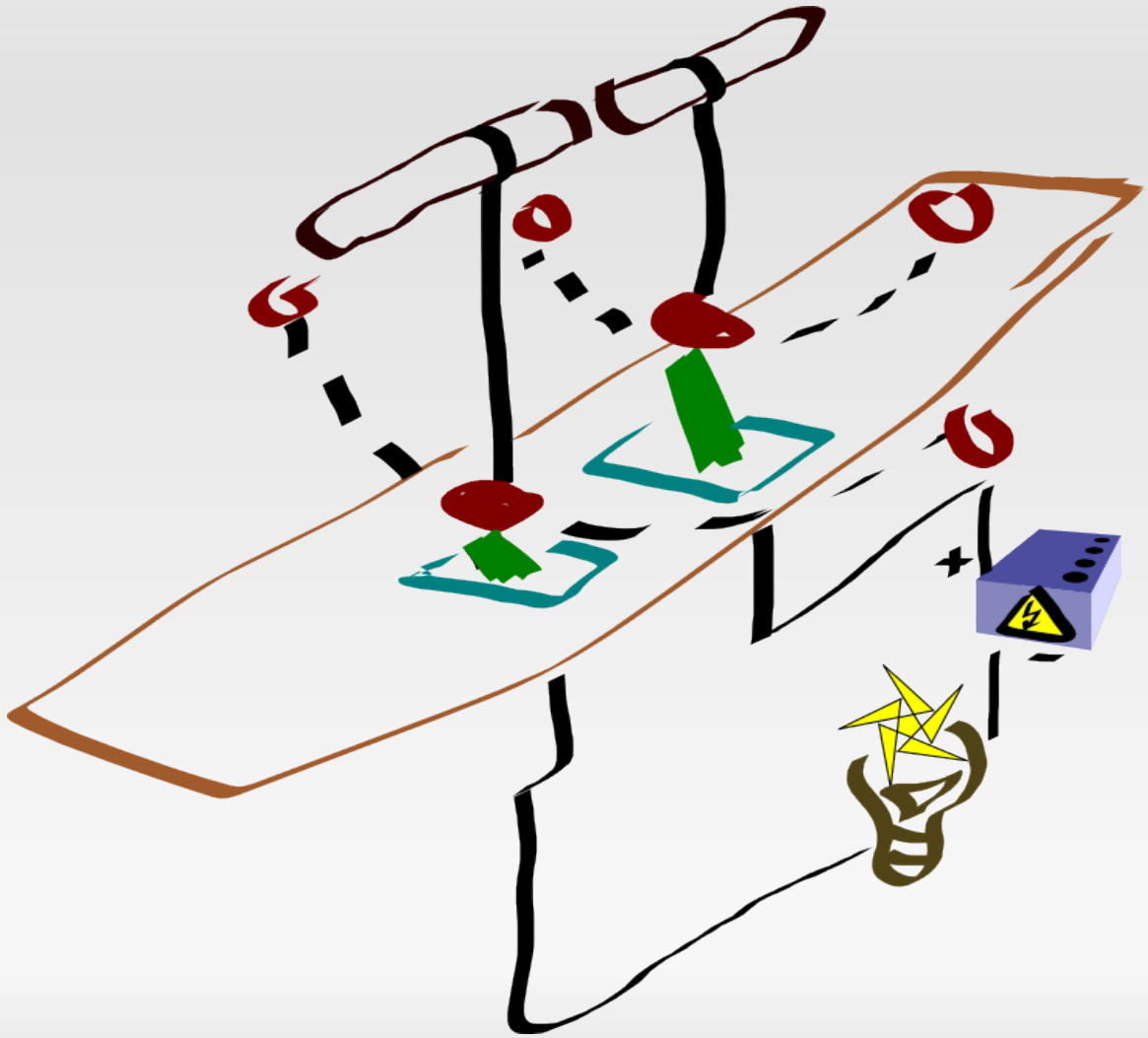


Base original
1/2 Oscillator response

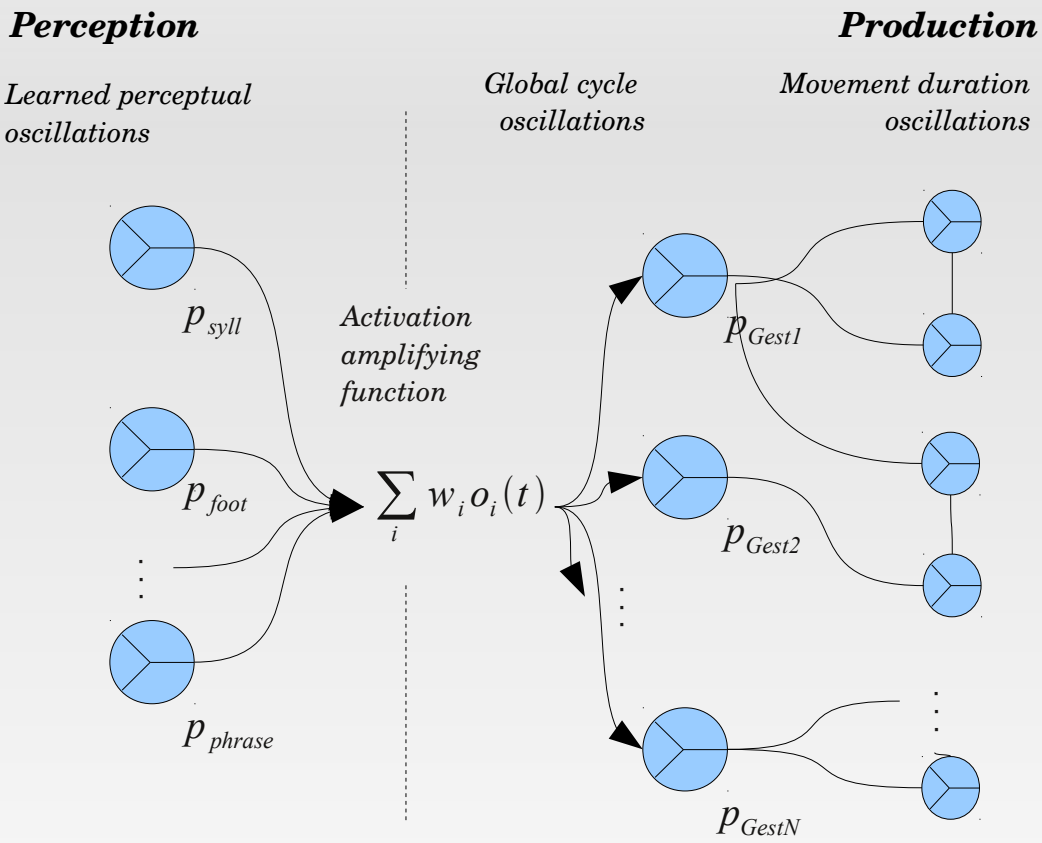
Dynamic perception-production oscillation model in human-machine communication

- Oscillators implemented in the brain, responsible for time perception, timing and coordination issues
- Evidence in psychology and phonetics for at least two oscillating mechanisms, one responsible for perception (=information clustering, rhythm perception, time windows etc) and one for production (motor coordination tasks)
- Back channel modeling in SFB673
- Modelling a timing mechanism in communication would provide more naturalness to the human-machine communication AND provide insights of possible implementations in the human brain

Basic idea



General model architecture



$$\vec{y}(t'') = S\{\sum_i w_i o_i(t')\}$$

- **Summed activation provoked by simultaneous firing**
- **Global cycle defines re-occurrence of output events**
- **The duration of an output event can be defined as an oscillation**

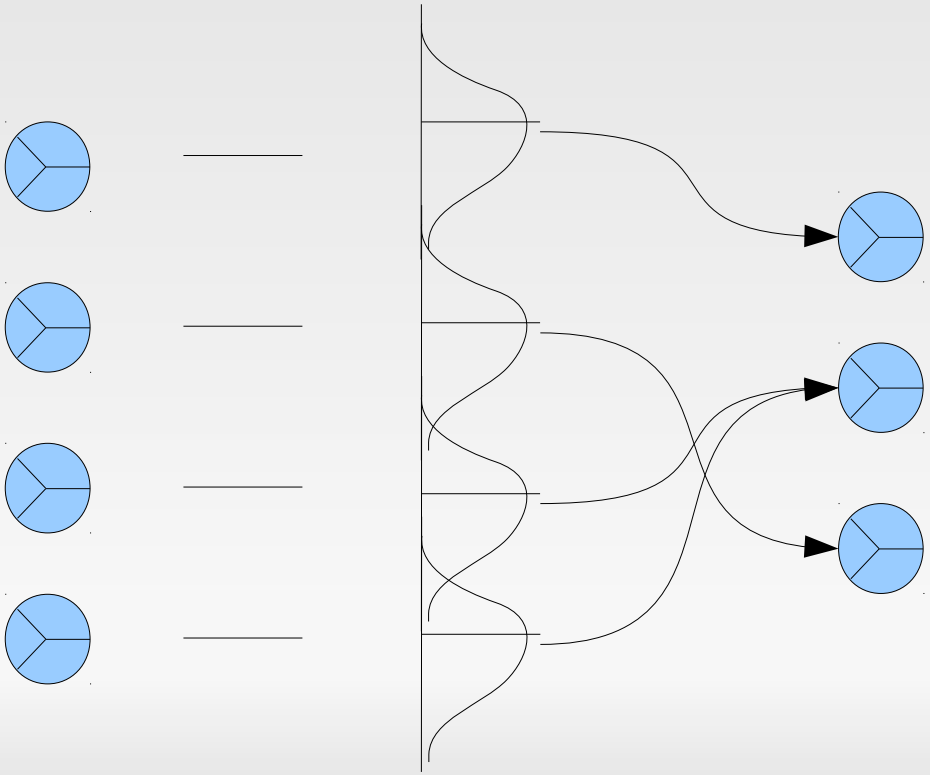
  → timing between the output events
- **Weights can control activation potential**

Experiments with the model

- First experiment to test the „percpetion“ part of the model and find the optimal parameter configuration
- Second experiment: letting Max, an artificial agent, make rhythmic moves to a rap song
- Results show the general capability of the model entrain to the rhythm and produce a rhythmic output
- Problems occur on higher rhythmic levels
  → probably those levels are not really rhythmic anymore

Future work: Hybrid model layout

- Real rhythms are rather *quasi-periodic* (if periodic at all)
- A metric level as a learned statistical distribution of pulses (motivated by exemplar theory, Pierrehumbert (2001)



- Mean values as input for upcoming layers
- Hybrid oscillator-statistical approach

GRÀCIES!