

# A mixed-lingual phonological component in polyglot TTS synthesis

Harald Romsdorfer, Beat Pfister and René Beutler

Speech Processing Group  
Computer Engineering and Networks Laboratory  
ETH Zurich, Switzerland

MLMI'04 workshop, Martigny

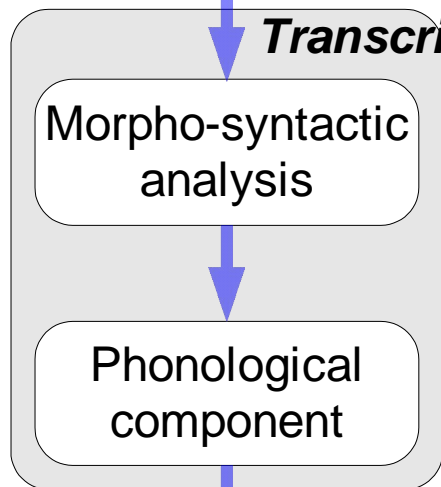
# Outline

- Overview of polyglot TTS system *polySVOX*
- Mixed-lingual phonological component
- Statistical models in prosody control
- Audio examples

# polySVOX overview

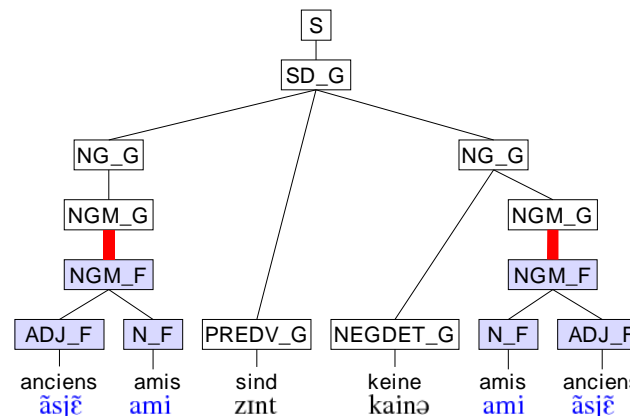
**Sentence**

**Transcription**



*Annotated syntax tree*

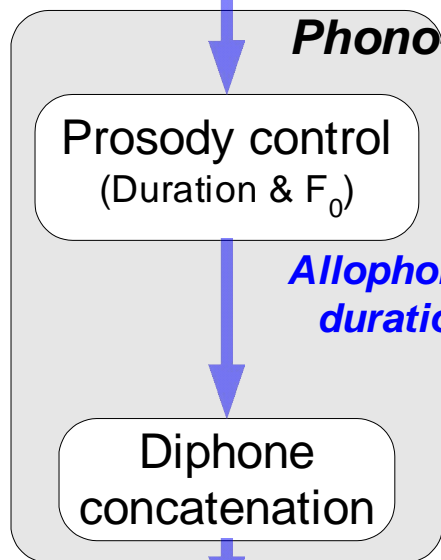
"Anciens Amis sind keine Amis anciens."  
("Ex-friends are no old friends.")



**Phonological representation**

(P) [1]\F\?ãs-jẽ- [3]za-mi- #{2}(T) \G\zɪnt- [2]k<sup>h</sup>ai-nə- [1]\F\?a-mi- [3]zãs-jẽ.

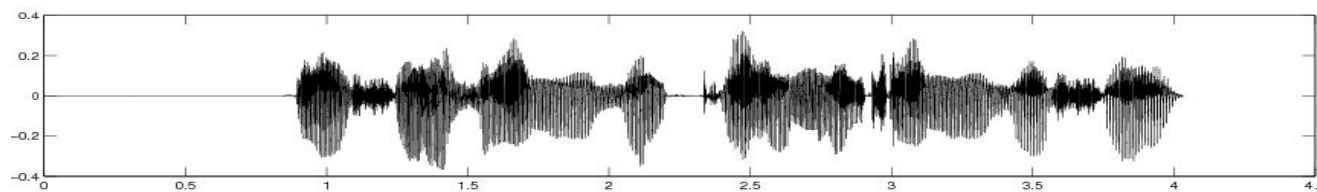
**Phono-acoustical model**



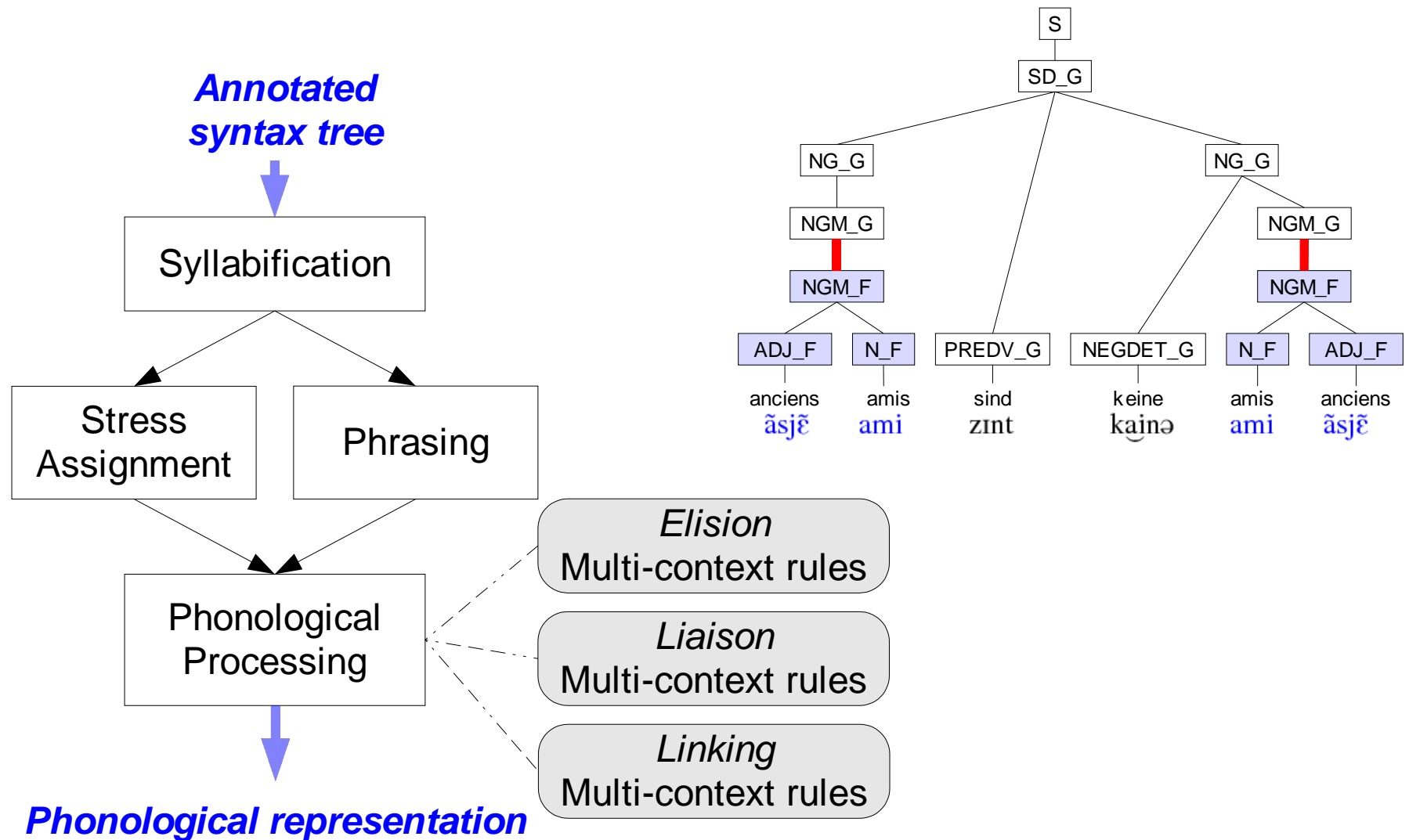
*Allophone sequence with duration and F<sub>0</sub> values*

\F\?	61	187	187	188	188	190
\F\~a	110	190	196	203	210	217
\F\s	104	217	221	223	228	243
\F\j	58	243	251	257	257	256
\F\~E	54	256	254	251	239	227
\F\z	83	227	210	201	201	201
.	.	.	.	.	.	.

**Speech signal**



# Mixed-lingual phonological component



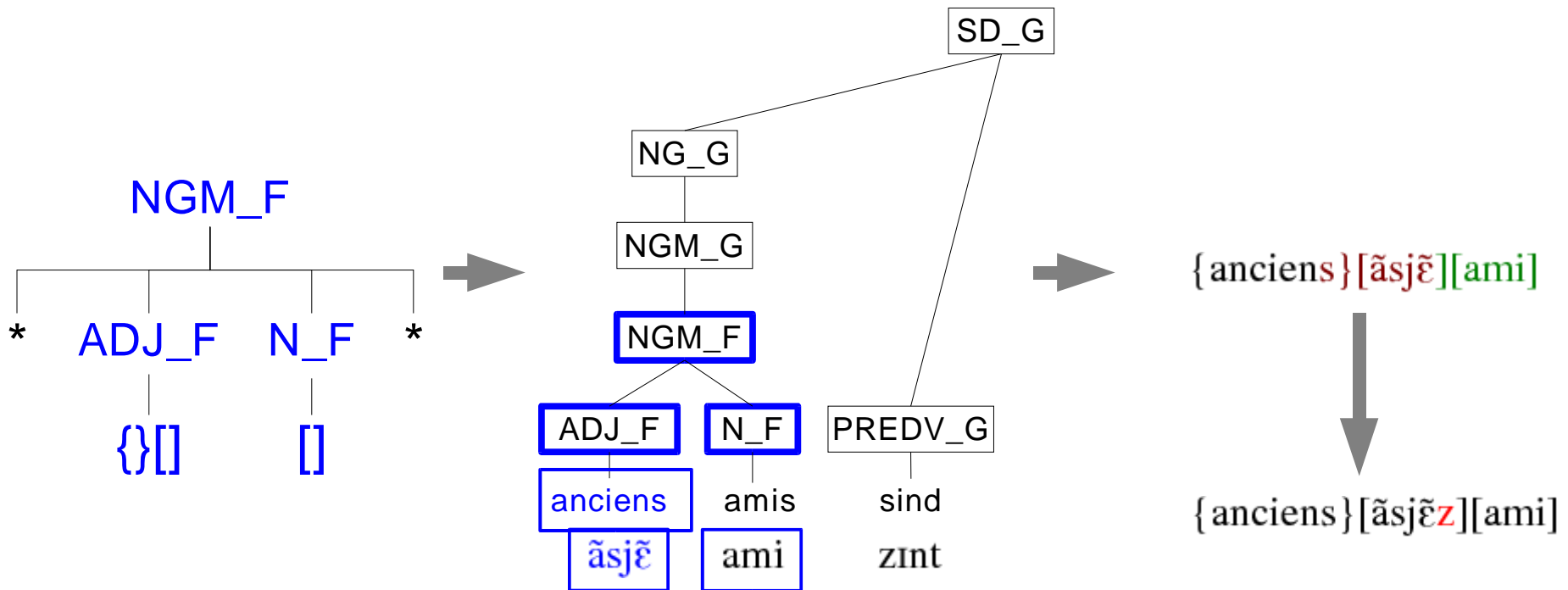
(P) [1]\F\?ãs-jẽ- [3]za-mi- #{2}(T) \G\zɪnt- [2]k<sup>h</sup>ai-nə- [1]\F\?a-mi- [3]zãs-jẽ.

# Multi-context rules

*subtree pattern* :  $X / Y \Leftrightarrow L \_ R ;$

Example: mandatory liaison within French noun groups between a noun and a preceding adjective:

$NGM\_F ( * ADJ\_F \{ \} [ ] N\_F [ ] * ) : @ / 'z' \Leftrightarrow 's' '}' '[' \{ \%P \} \%V \_ ']' '[' \%V ;$



# Multi-context rules: examples

*French mandatory liaison* between a personal pronoun and a conjugated verb or "en" or "y" as a pronoun or adverb:

```
PERS_F {}[] PRONADV_F [] : @ / 'z' <=> 's' '}' '[' {%P} %V _ ']' '[' %V ;
```

"vous en reprenez"  
("you take it away")

(T) \F\vu-z ã- RƏ-prƏ-[1]ne.

"Où partez vous en vacances?"  
("Where do you go on vacation?")

(T) \F\u- pɑR-te- vu- ã- va-[1]kãs.

*English linking [r]* is pronounced only when a word ends graphemically on "r" and the next word starts with a vowel:

```
*_E {}[] *_E [] : @ / 'r' <=> 'r' '}' '[' {%P} _ ']' '[' %V ;
```

"four eggs"

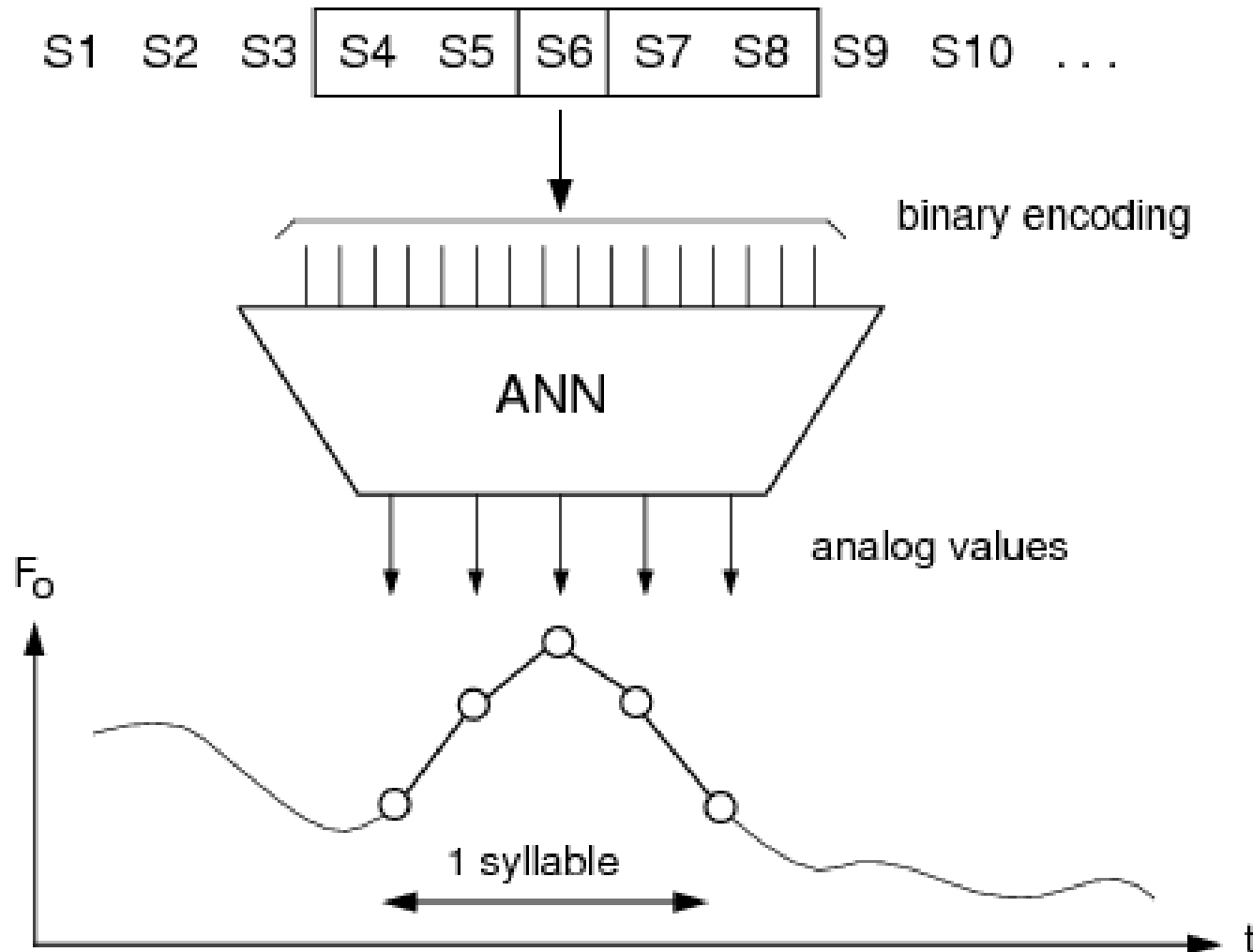
(T) \E\fɔ:r- [1]egz.

"four dogs"

(T) \E\fɔ:- [1]dɒgz.

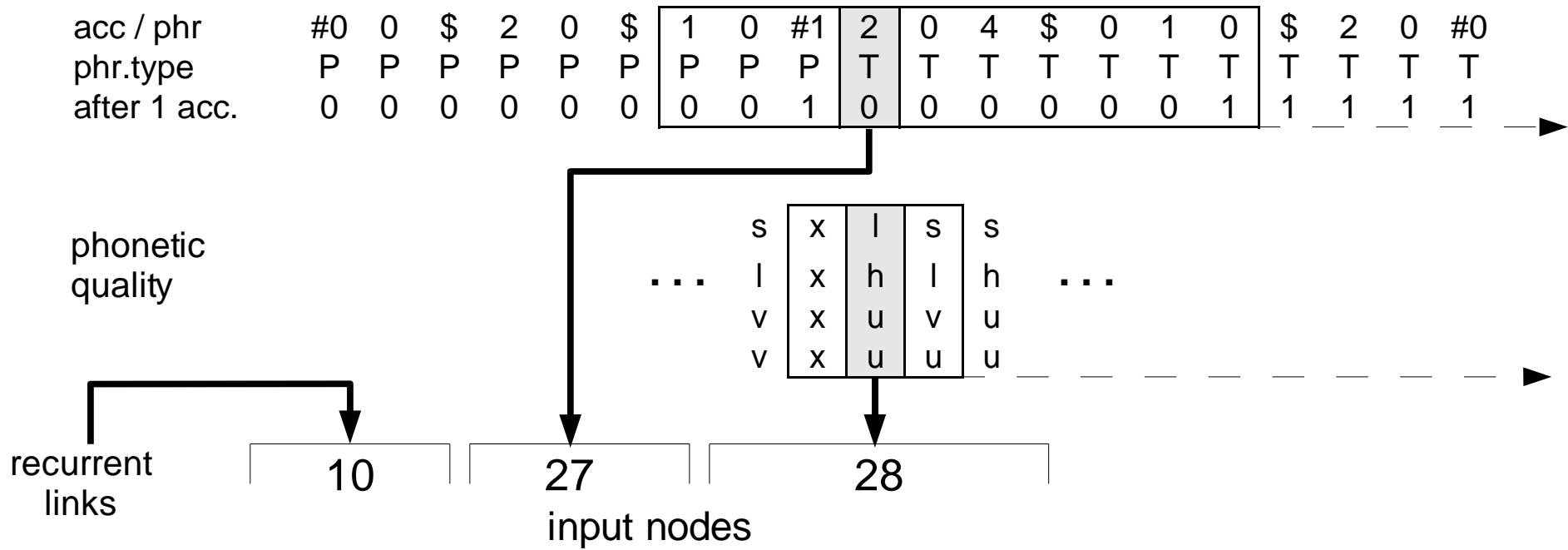
# Prosody control

(P) [1]\F\ʔã̃s-jẽ- [3]za-mi- # {2}(T) \G\zɪnt- [2]k<sup>h</sup>ai-nə- [1]\F\ʔa-mi- [3]zã̃s-jẽ.



# Prosody control

## Input encoding:



## Architecture of neural net:

- Recurrent neural network with 2 hidden layers
- 65 input nodes
- 10 recurrent links from 2. hidden layer
- 5 output nodes for syllable pitch values



# Audio examples

German

Polyglot

- ■ ■ – **Anciens Amis** sind keine **Amis anciens**.  
(*Ex-friends are no old friends.*)
- ■ ■ – Die grossspurige **Greatest Nation** hat die kulturbewusste **Grande Nation** verdrängt.  
(*The cocky Greatest Nation has displaced the culture-conscious Grande Nation.*)
- ■ ■ – Ein **Cheminée**ofen ist kein **Must**, sondern bloss **nice to have**.  
(*A fireplace is not a must, but just nice to have.*)
- ■ ■ – **Lifestyle** outet sich oft als **niveaulose Eventkultur**.  
(*Lifestyle often emerge as dull event culture.*)
- ■ ■ – Mein **Cousin** surft im **World Wide Web**.  
(*My cousin surfs the world wide web.*)