

## **Einsatz von textgebundener Sprachsynthese**

- „black-box“ Text-to-Speech (TTS) für verschiedene Sprachen
- für ein spezifisches Einsatzgebiet (Telefonnummernsynthese, Perzeptionstests, ...)
- als Sprachsynthese-Entwicklungsumgebung (Entwicklung verschiedener u.U. sprachabhängiger Module, linguistische Forschung)

# *Text-to-speech*

---

Hr. Müller, kommt er schon um 11:45 h?

## Text-to-speech

---

Hr. Müller, kommt er schon um 11:45 h?

Herr Müller<sub>Komma</sub> kommt er schon um elf Uhr fünfundvierzig<sub>Fragezeichen</sub>

## Text-to-speech

---

Hr. Müller, kommt er schon um 11:45 h?

Herr Müller<sub>Komma</sub> kommt er schon um elf Uhr fünfundvierzig<sub>Fragezeichen</sub>

\_hER mY16 kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC\_

## Text-to-speech

---

Hr. Müller, kommt er schon um 11:45 h?

Herr Müller<sub>Komma</sub> kommt er schon um elf Uhr fünfundvierzig<sub>Fragezeichen</sub>

\_hER mY16 kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC\_

[[\_hER mY16] [kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC\_]]

# Text-to-speech

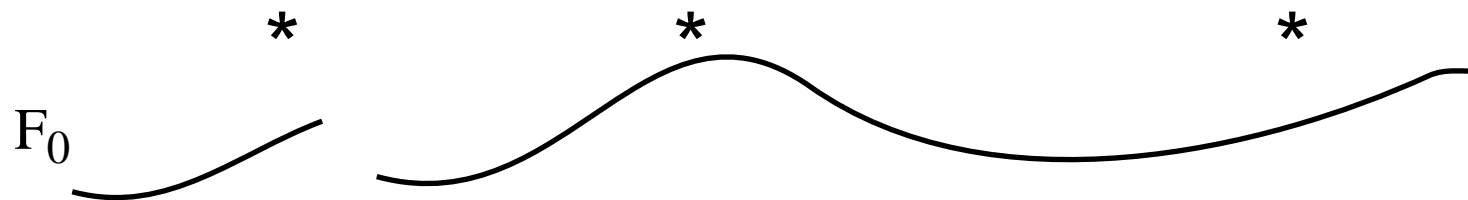
---

Hr. Müller, kommt er schon um 11:45 h?

Herr Müller<sub>Komma</sub> kommt er schon um elf Uhr fünfundvierzig<sub>Fragezeichen</sub>

`_hER mY16 kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC_`

`[[_hER mY16] [kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC_]]`



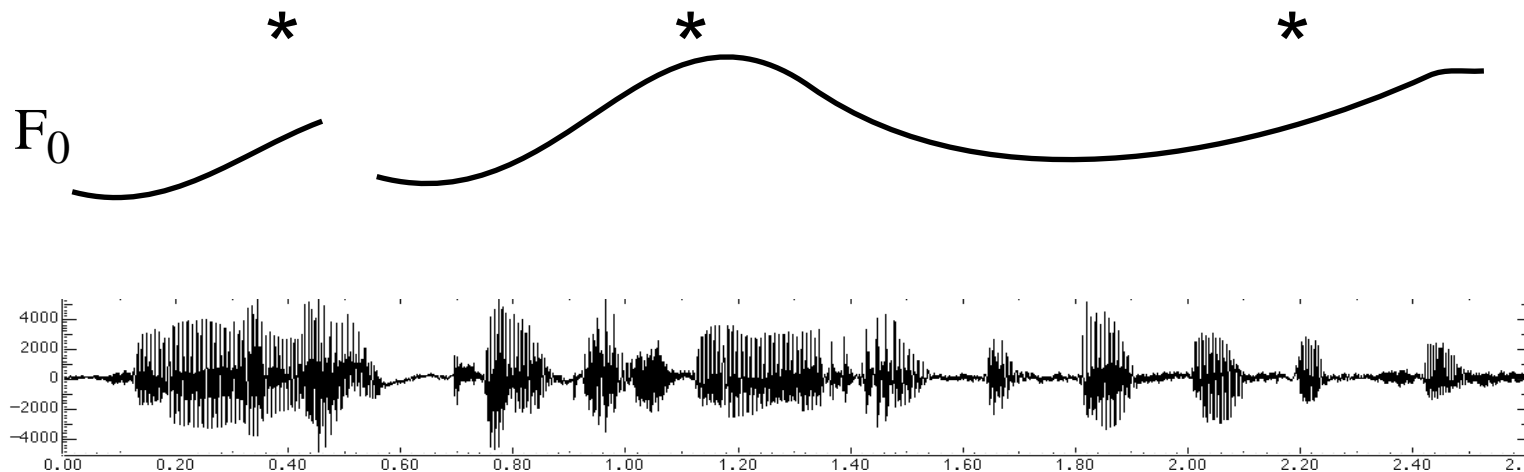
# Text-to-speech

Hr. Müller, kommt er schon um 11:45 h?

Herr Müller<sub>Komma</sub> kommt er schon um elf Uhr fünfundvierzig<sub>Fragezeichen</sub>

`_hER mY16 kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC_`

`[[_hER mY16] [kOmt e:6 So:n ?Um ?Elftu:6 fYnf?UntfIRtsIC_]]`



## **Programmfluß / Datenstruktur**

- Pipeline zwischen Modulen
- globale "Utterance Structure", auf die alle Module zugreifen

# *Sprachsynthesesystem Festival*

---

## ***Verschiedene Ebenen der Interaktion mit Festival:***

- Unix-Shell,  
Einbindung in eine anderes Programm
- Scheme-Scriptsprache,  
konfigurierbare Vorverarbeitung, Intonation, ...
- C++  
Neuentwicklung, Benutzung der *Speech-Tools* library

## **Eigenschaften von Festival:**

- Scriptsprache auf der Grundlage von Scheme (Interpreter-basiert)
- Module in C/C++ (Compiler-basiert)
- Gemeinsame zugrundeliegende Datenstruktur (*Utterance* mit *streams*), Zugriff auf Scheme- und C++ - Ebene
- statistische Standardwerkzeuge (CART, Viterbi, ...)

# *Sprachsynthesesystem Festival*

---

## ***Festival 1.4***

- Komplettes englisches und spanisches TTS
- Interface to MBROLA-Diphonen (weitere Sprachen!)
- Konfigurierbare sprachabhängige Module (Phonemsatz, Lexika, Intonation, ...)
- Plattformen: Unix, Linux, WindowsNT/95 (als vorläufige Version)
- APIs, Beispielanwendungen (`saytime`), client-server, ...
- XML-Interface

## ***utterance und relations***

- die grundlegende Einheit ist ein *item* (z.B. ein Phon)
- *items* können *features* besitzen.
- items werden in Listen organisiert (*relations*, z.B. eine Phonemkette)
- zwischen den items können hierarchische Strukturen gebildet werden (z.B. Syntaxbaum)
- *utterance*: Übergeordnete Datenstruktur von Festival.

# Sprachsynthesystem Festival

---

```
(defuttType Text
  (Initialize utt)
  (Text utt)
  (Token utt)
  (POS utt)
  (Phrasify utt)
  (Word utt)
  (Pauses utt)
  (Intonation utt)
  (PostLex utt)
  (Duration utt)
  (Int_Targets utt)
  (Wave_Synth utt))
```