

Speech and Language Technology for Linguists and Other Human Scientists

Interspeech Tutorial T-S1-R3

Daniel Hirst

Laboratoire Parole et Langage, CNRS and Université de Provence
daniel.hirst@lpl-aix.fr

Makuhari, 2010 September 26

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Linguists as human scientists...

Linguistics needs to become a science.

Science is

- ▶ predictive
- ▶ cumulative
- ▶ empirically testable

What is science?

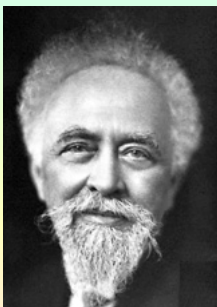


Figure: Jean Baptiste Perrin (1870-1942).

scientific method: explaining visible complexity
by invisible simplicity.
(expliquer le visible compliqué par l'invisible simple.)

What is science?

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives



Figure: William of Ockham (1285-1349).

Ockham's razor: Don't use unnecessary variables
(*Entia non sunt multiplicanda praeter necessitatem*)

What is science?



Figure: Jorma Rissanen (b. 1932)

MDL: Minimum Description Length

The best hypothesis for a given set of data is the one that leads to the best compression of the data.

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Getting a corpus

Making recordings

Manipulating text

Aligning sound and text

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Getting a corpus

Making recordings

Manipulating text

Aligning sound and text

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Getting a corpus

Maybe there is already a corpus you can use?

Eurom1 Includes 40 4-sentence passages
(aka MULTEXT)

ELRA/ELDA Europe

LDC USA

VOXFORGE <http://www.voxforge.org>

Other Google it...

Making recordings

- ▶ Anything is better than nothing
- ▶ Good is better than bad
- ▶ Standard is better than non-standard
- ▶ Legal problems

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Getting a corpus

Making recordings

Manipulating text

Aligning sound and text

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Manipulating text

Speech Technology
for Human Scientists
12/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Getting a corpus

Making recordings

Manipulating text

Aligning sound and text

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Tasks we need to be able to do

- ▶ Clean up text
- ▶ Convert orthographic to phonetic transcription
- ▶ Manipulate file structures
- ▶ Learn to use Regular Expressions (regex)

Aligning sound and text

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Getting a corpus

Making recordings

Manipulating text

Aligning sound and text

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Festival <http://www.cstr.ed.ac.uk/projects/festival>

Festvox <http://festvox.org>

HTS <http://hts.sp.nitech.ac.jp>

Easyalign <http://latlcui.unige.ch/phonetique>

WinPitch <http://www.winpitch.com>

Jtrans <http://www.loria.fr/~cerisara/jtrans/index.html>

P2FA <http://www.ling.upenn.edu/phonetics/p2fa>

Praat <http://www.praat.org>

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Explicit predictive models

Possible examples: prosodic typology and structure

- ▶ lexical typology:
 - ▶ tone / quantity / stress / pitch accent
- ▶ rhythmic typology:
 - ▶ stress-timed / syllable-timed / mora-timed
- ▶ tonal typology
 - ▶ compressing / truncating
 - ▶ left-headed / right-headed
- ▶ prosodic structure
 - ▶ syllable / stress group
 - ▶ intermediate phrase / intonational phrase
 - ▶ tonal unit / rhythm unit

Testing these models with empirical data

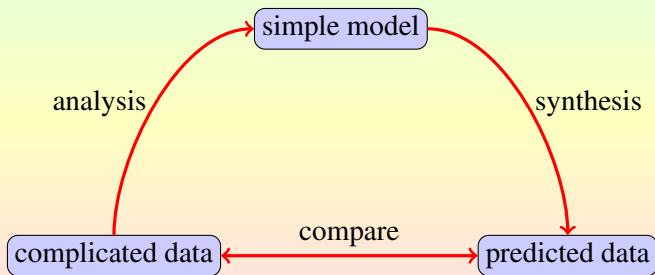


Figure: The Analysis by Synthesis paradigm

Statistics with R

Speech Technology
for Human Scientists
17/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

The R project <http://www.r-project.org/>

Baayen 2008 Analyzing Linguistic Data.

<http://www.ualberta.ca/baayen/publications.html>

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

**Modelling speech
rhythm**

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Rhythmic typology

From data to models

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Rhythmic typology

Two types of rhythm

Arthur Lloyd James 1940, Kenneth Pike 1945

<http://www.britishpathe.com/record.php?id=82205>

- ▶ Machine gun rhythm
- ▶ -----
- ▶ syllable timed: regular syllable/vowel onsets
- ▶ Morse code rhythm
- ▶ . — . . — . — . .
- ▶ stress timed: regular stressed syllable/vowel onsets

Typological distinction

Abercrombie 1967

- ▶ Stress timed (English, Russian, Arabic)
- ▶ Syllable timed (French, Telugu, Yoruba)

Ladefoged 1975

- ▶ Mora timed (Japanese, Tamil)

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

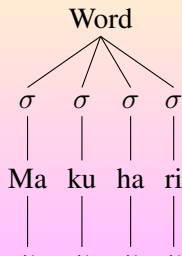
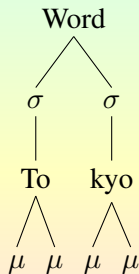
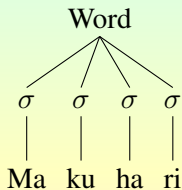
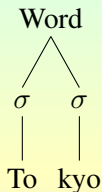
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Moras (or morae)

mora “something of which a short syllable has only one but a long syllable has two” J. McCawley 1968



Syntactic structure

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

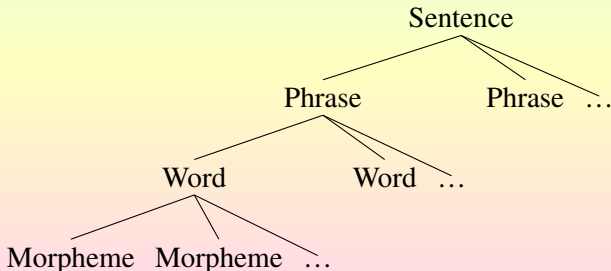
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

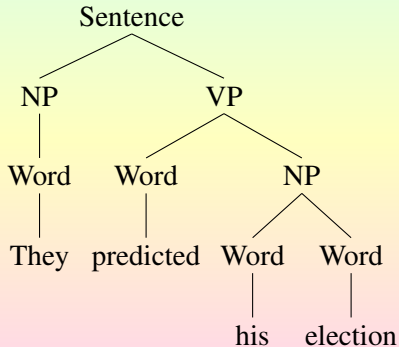
Categories Sentence > Phrase* > Word > Morpheme

* = recursive category



Syntactic structure

They predicted his election



Phonological structure

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

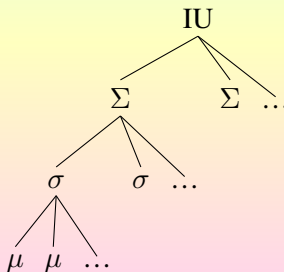
From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

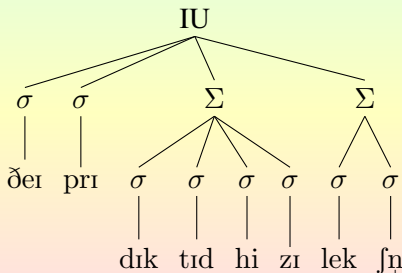
Perspectives

Categories Intonation Unit (IU) > Stress group (Σ)
> Syllable (σ) > Mora (μ)
no recursive categories (?)



Phonological structure

They predicted his election



Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

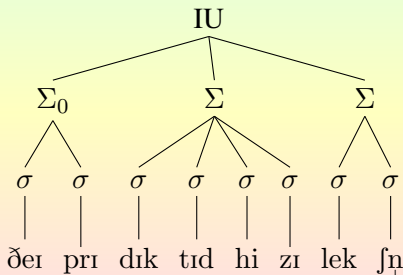
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Phonological structure

They predicted his election



Experimental evidence for rhythmic typology

Compare *mean* and *sd* for σ and Σ

Roach 1982 Two minutes of spontaneous speech

6 languages Σ : [EN, RU, AR]* σ : [FR, TE, YO]

Prediction smaller *sd* for Σ or σ

Result No significant difference

Dauer 1983 5 languages [EN, IT, EL, ES, TH]

Conclusion Differences are structural

- simple/complex consonant clusters
- short vowels/long vowels/diphthongs
- \pm vowel reduction

*ISO 639-1/2: 2/3 letter language codes

Psycholinguistics

Speech Technology
for Human Scientists
29/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Mehler et al. 1996 The TIGRE

TIGRE Text Intensity Grid REpresentation

Bertoncini et al 1995, Nazzi et al. 1997 French neonates

discriminate NL vs JA or EN vs IT

but not EN vs NL or IT vs ES

The TIGRE

Consonantal and vocalic intervals

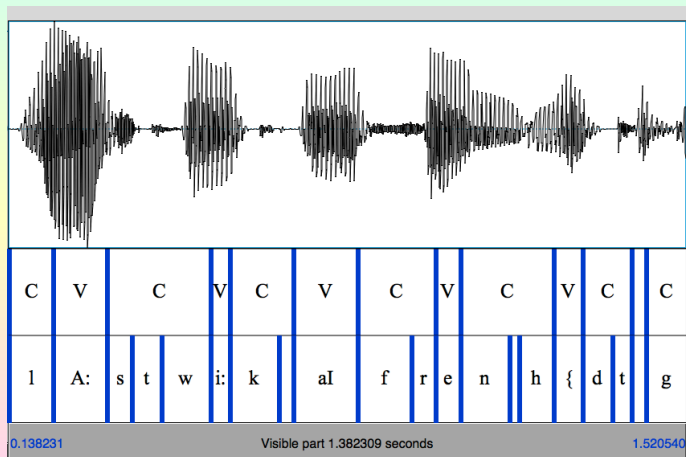


Figure: C and V intervals "Last week my friend had to go..."

A new metric

$$\Delta C \quad sd(duration_C)$$
$$\%V \quad 100 \cdot \frac{\sum duration_V}{\sum duration_V + \sum duration_C}$$

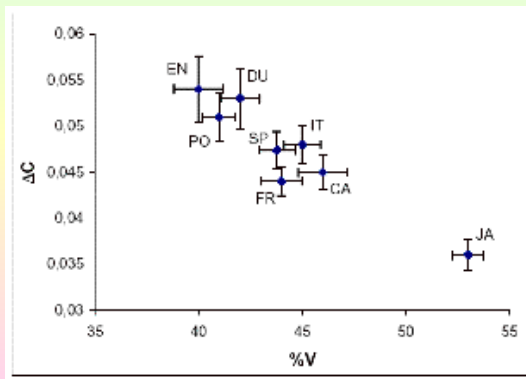


Figure: A new metric: Ramus 1999

Rhythm of Speech and Text

Speech Technology
for Human Scientists
32/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

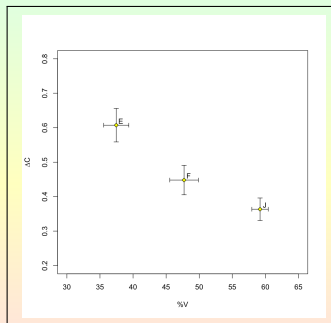
Rhythmic typology

From data to models

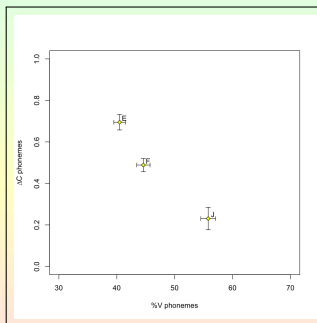
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives



a. From acoustic data
(Speech)



b. From phoneme count
(Text)

Speech and Text

Correlation of durations and phoneme counts

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

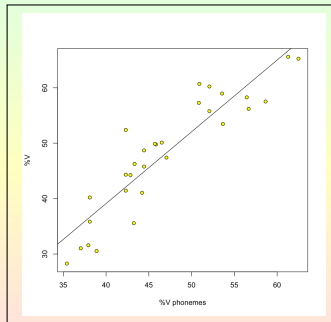
Rhythmic typology

From data to models

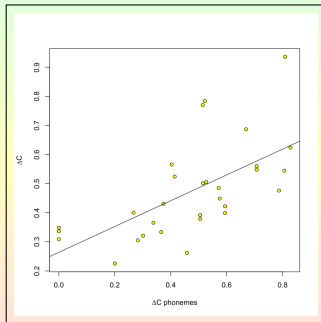
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives



a. V: text/speech
 $r = 0.911$



b. C: text/speech
 $r = 0.627$

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Rhythmic typology

From data to models

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Klatt's 'unsolved problem'

One of the unsolved problems in the development of rule systems for speech timing is the size of the unit (segment, onset/rhyme, syllable, word) best employed to capture various timing phenomena.

Klatt (1987) p.760

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

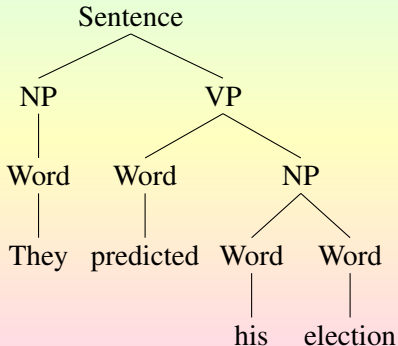
Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

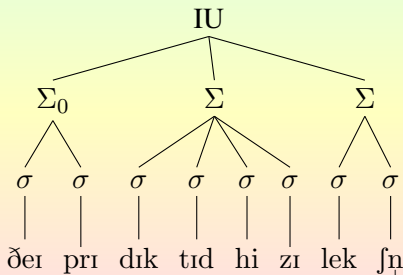
Syntactic structure

They predicted his election



Phonological structure

They predicted his election



A forgotten model

Jassem 1950 proposed **distinct** models for tonal and for rhythmic structure

Intonation Unit aka Intontional phrase, prosodic phrase, Tone Group, etc.

Tonal Unit τ (like Σ): one stressed syllable (σ) plus any number of unstressed σ s

Rhythm Unit ρ does not cross word boundaries
summer dresses \neq some addresses
 $/\text{'s}\Lambda\text{m}\text{ə 'dres}\text{ɪ}\text{z}/ \neq /s\Lambda\text{m ə'dres}\text{ɪ}\text{z}/$

Jassem 1950

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

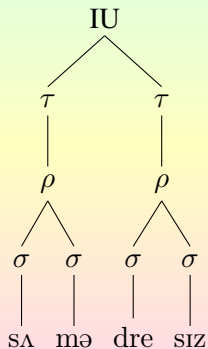
From data to models

Modelling speech
melody

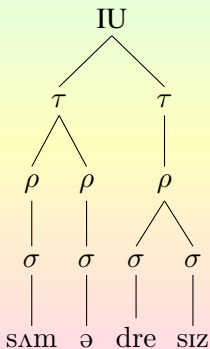
Speech synthesis and
re-synthesis

Perspectives

Summer dresses

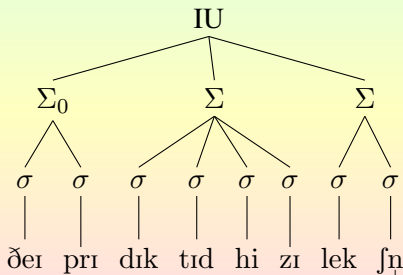


Some addresses



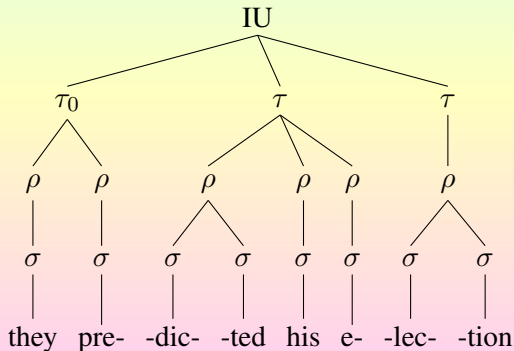
Phonological structure

They predicted his election



Phonological structure (Jassem)

They predicted his election



Rhythmic typology

Two types of rhythm

Arthur Lloyd James 1940, Kenneth Pike 1945

<http://www.britishpathe.com/record.php?id=82205>

- ▶ Machine gun rhythm
- ▶ -----
- ▶ syllable timed: regular syllable/vowel onsets
- ▶ Morse code rhythm
- ▶ . — . . — . — . .
- ▶ stress timed: regular stressed syllable/vowel onsets

Linear model

Faure, Hirst & Chafcouloff 1980 $d_{\Sigma} = 0.080 + 0.14 \cdot n_{\sigma}$

Eriksson 1991 - ES, GR, IT $d_{\Sigma} = 0.100 + 0.1 \cdot n_{\sigma}$

ibid - EN, SV, IS $d_{\Sigma} = 0.200 + 0.1 \cdot n_{\sigma}$

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

The Aix-Marsec Database

SEC Spoken English Corpus. Knowles et al. 1996
London-Lancaster Corpus. 5.5 hours of
“authentic” speech
68 speakers, c 50000 words
prosodic markup - Tonetic Stress Marks
(Knowles & Williams)

MARSEC Machine-Readable SEC. Roach et al. 1993
words aligned with signal

Aix-Marsec Database Auran, Bouzon & Hirst 2004
Hirst, De Looze, Auran & Bouzon forthcoming.
phonetic transcription aligned with signal
Prosodic Structure (Praat TextGrids)

TextGrid from Aix-Marsec

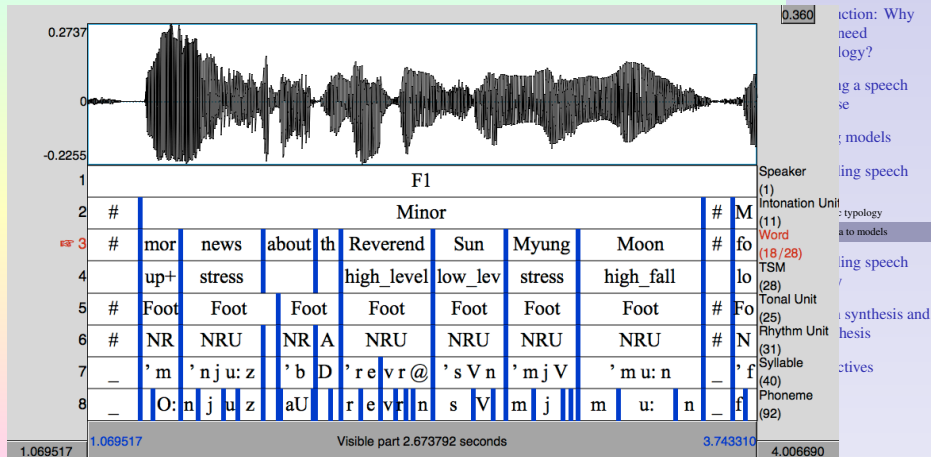


Figure: “More news about the Reverend Sun Mjung Moon”

Foot vs number of syllables

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

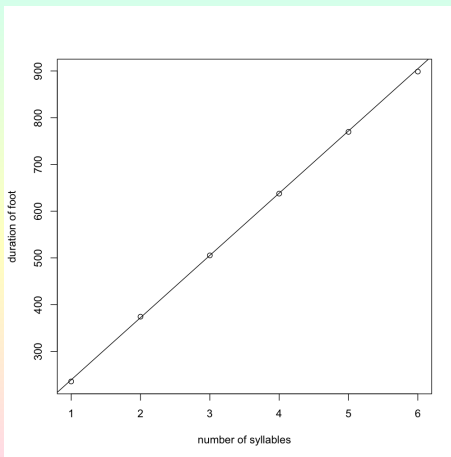


Figure: Duration of foot (Σ) as function of number of constituent syllables. Data from Aix-Marsec database cf Hirst & Bouzon 2005.

Stressed vs unstressed syllables

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

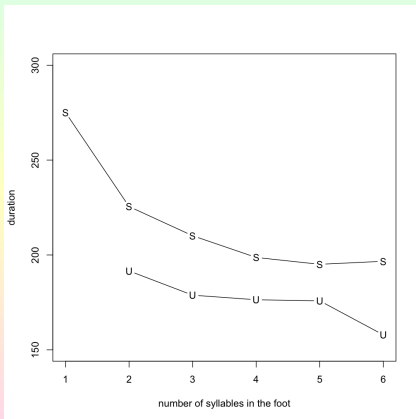


Figure: Duration of stressed (s) and unstressed syllables (u) as function of number of syllables in the foot ($= \Sigma$). Data from Aix-Marsec database. cf Hirst & Bouzon 2005.

Phoneme duration

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

- ▶ Most significant factor is phoneme identity
- ▶ Neutralise this using z-score
- ▶
$$z = \frac{d_i - \text{mean}_{d/p}}{sd_{d/p}}$$

Z-score of phonemes in (ρ)

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

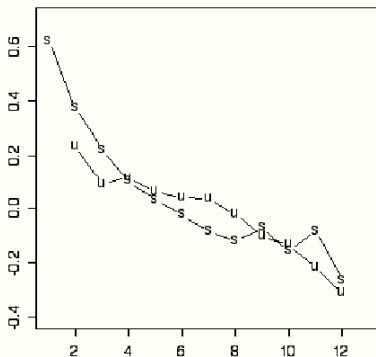


Figure: z-score as function of number of phonemes in narrow rhythm unit (ρ). Data from Aix-Marsec database. cf Hirst & Bouzon 2005.

Intonation Unit Final lengthening

Speech Technology
for Human Scientists
50/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

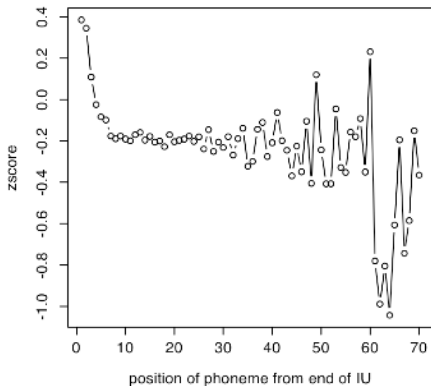


Figure: Intonation Unit Final lengthening

Z-score of phonemes in ρ

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

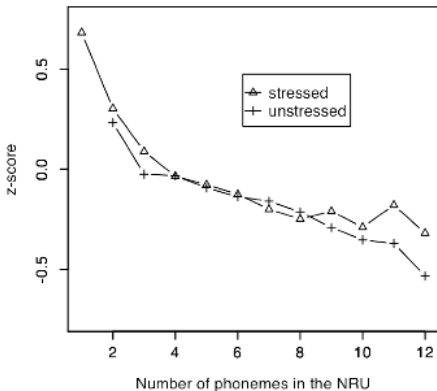


Figure: z-score as function of number of phonemes - excluding utterance final phonemes. Data from Aix-Marsec database. cf Hirst & Bouzon 2005.

Results

Negative correlation z-score : size of σ (ns) $< \text{Word} < \Sigma < \rho$

No specific effect stressed vs unstressed syllable

A simple linear model?

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Rhythmic typology

From data to models

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

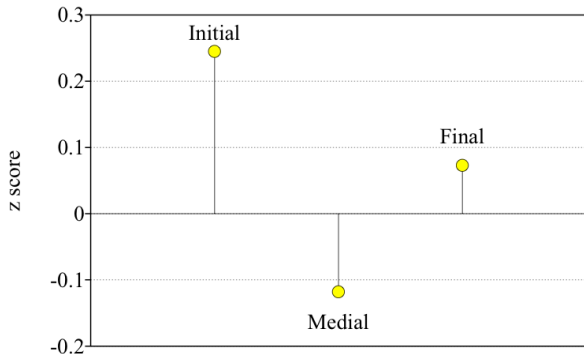


Figure: Significant differences of phoneme duration. Data from Aix-Marsec database. cf Hirst 2009.

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and re-synthesis

Perspectives

Analysing f_0

Speech Technology
for Human Scientists
56/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Extract from Aix-Marsec corpus. (Passage A01-01)

Good morning. More news about the Reverend Sun
Myung Moon, founder of the Unification Church, who's
currently in jail for tax-evasion.

Detecting f_0

Passage A01-01

Speech Technology
for Human Scientists
57/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

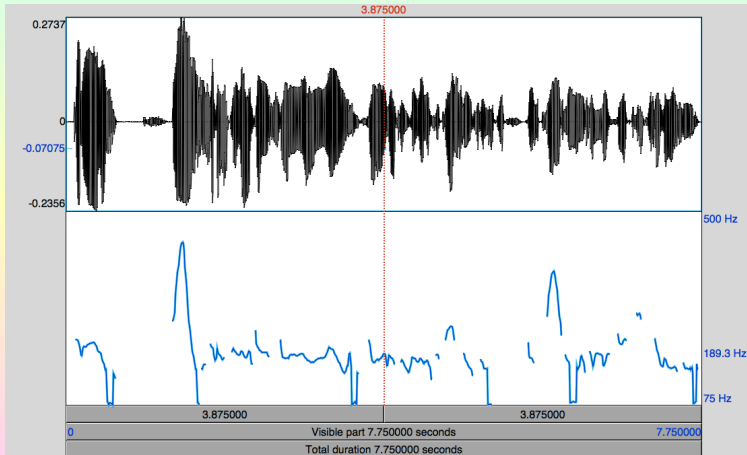


Figure: Praat's default parameters for F0. Range = [75-500]

Detecting f_0

Passage A01-01

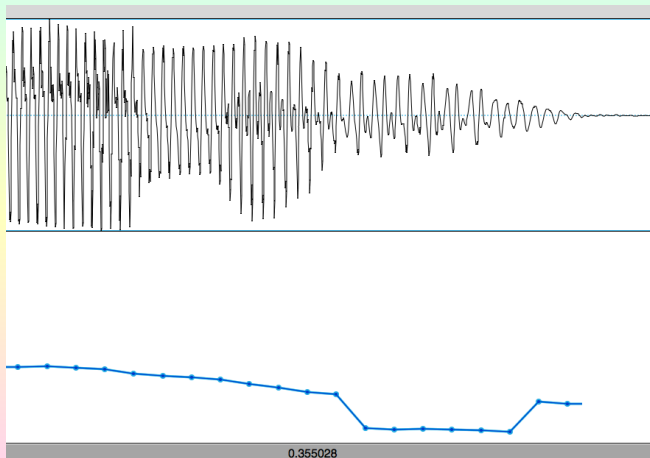


Figure: Example of octave error: pitch halving

The two-pass method

- ▶ First pass: Default parameters:
[50, 500]
- ▶ Calculate quantiles: q_{25} and q_{75}
- ▶ Second pass: Derived parameters:
[$0.75 * q_{25}$, $1.5 * q_{75}$]

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Detecting f_0

Passage A01-01

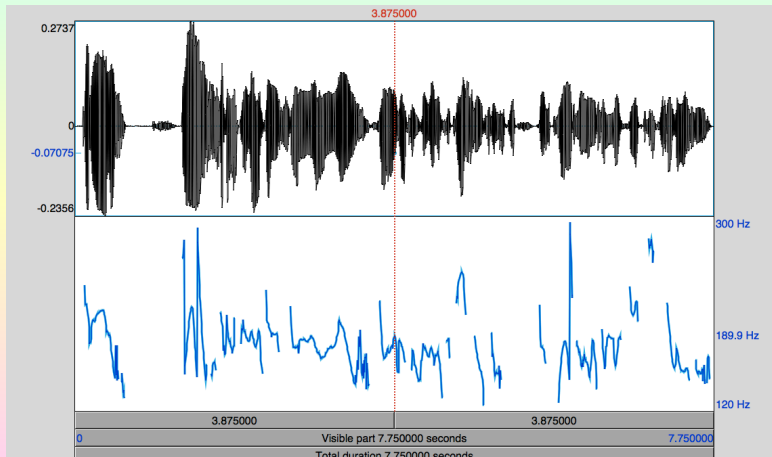


Figure: Parameters from the two-pass method. Range = [120-300]

Detecting f_0

Passage A01-01

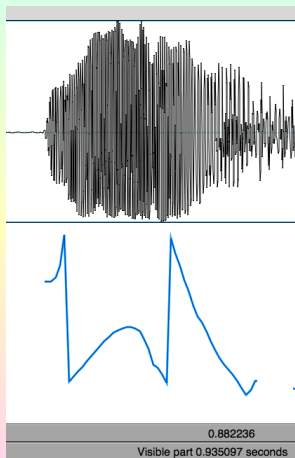


Figure: Example of octave error: pitch halving

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Detecting f_0

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

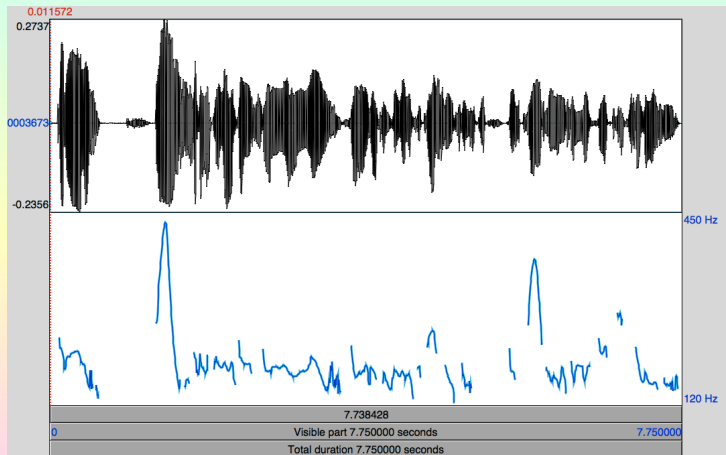


Figure: Adjusting for expressive voice. Range = [120-450]

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Problem for modelling f_0

"More news about the Reverend Sun Myung M..."

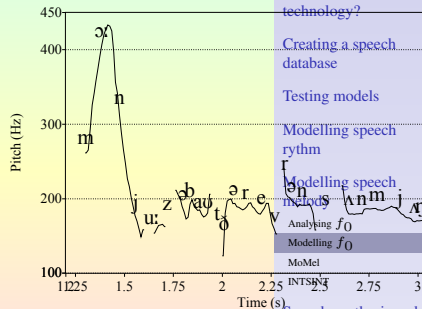
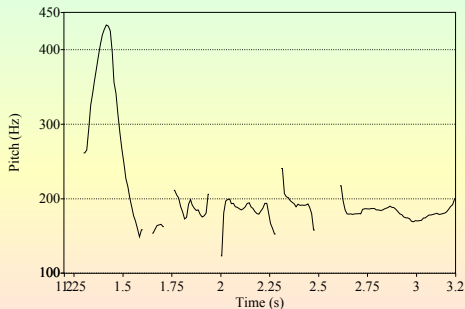


Figure: Two second extract of f_0 curve

- ▶ Raw F0 is discontinuous and not smooth.
- ▶ Here beginning and end is continuous and smooth
- ▶ Discontinuity is due to microprosodic effect of consonants

Synthesising intonation

Speech Technology
for Human Scientists
65/109

Daniel Hirst

“The end of the table that is furthest from you.”

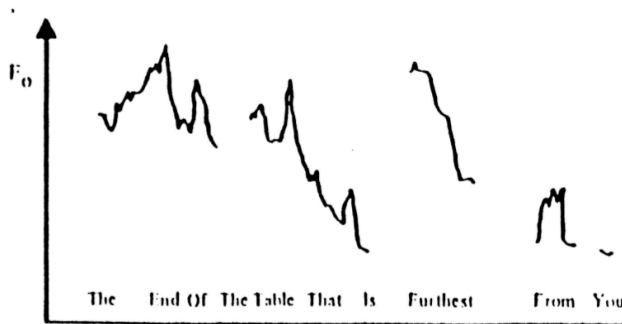


Figure: Example from Kloker (1975)

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Synthesising intonation

Speech Technology
for Human Scientists
66/109

Daniel Hirst

“The end of the table that is furthest from you.”

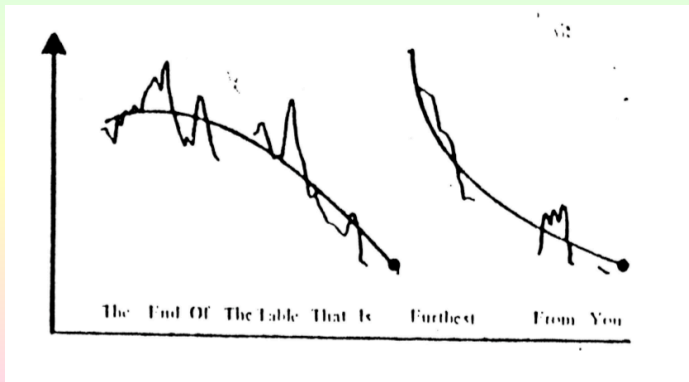


Figure: Gamma function: $y = at^b e^{ct}$

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

A lesson

Kloker's model relies on the discontinuity of the voiceless consonant /f/ in the word "furthest"

If the sentence had been "The end of the table that is nearest to you" with a sonorant consonant /n/ then this model would not have been appropriate.

Hirst's law An acoustic model should not depend on which end of the table you are talking about.

Using sonorants to illustrate intonation patterns

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

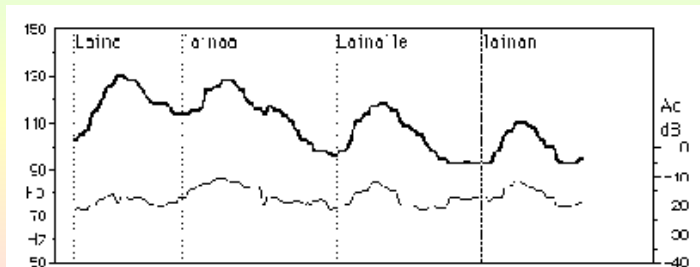


Figure: Finnish intonation (from Iivonen 1998)

Macro- and Micro-melody

Statement intonation

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

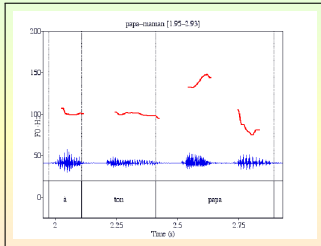
Modelling f_0

MoMel

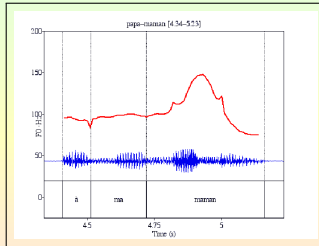
INTSINT

Speech synthesis and
re-synthesis

Perspectives



a. A ton papa.
/at̃papa/



b. A ma maman.
/amamamã/

Macro- and Micro-melody

Question intonation

Speech Technology
for Human Scientists
70/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

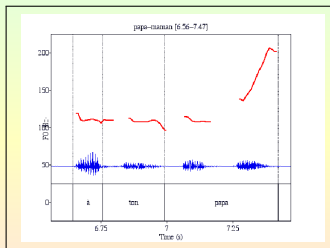
Modelling f_0

MoMel

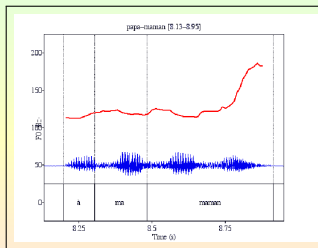
INTSINT

Speech synthesis and
re-synthesis

Perspectives



a. A ton papa ?
/at̃papa/



b. A ma maman ?
/amamamã/

General model for F0

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Raw f_0 is the combination of two components

- ▶ Macromelodic component: smooth and continuous
(Underlying intonation pattern)
- ▶ Micromelodic component: discontinuous
(Surface effect of phonemes)

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

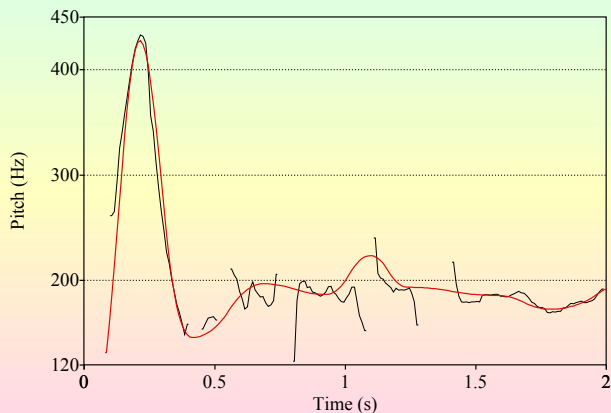


Figure: Macromelodic profile for extract from A01-01

Macromelodic and Micromelodic profiles

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

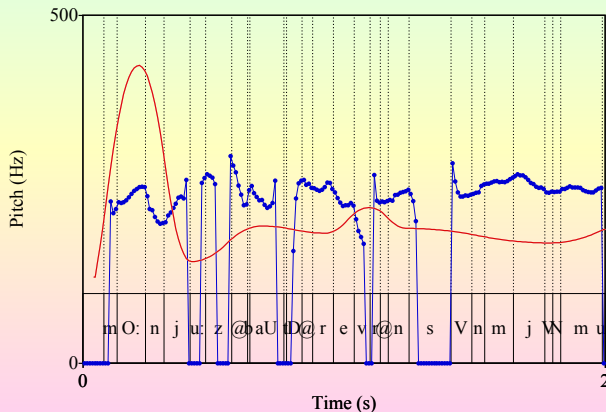


Figure: Macromelodic (red) and micromelodic (blue) profiles

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

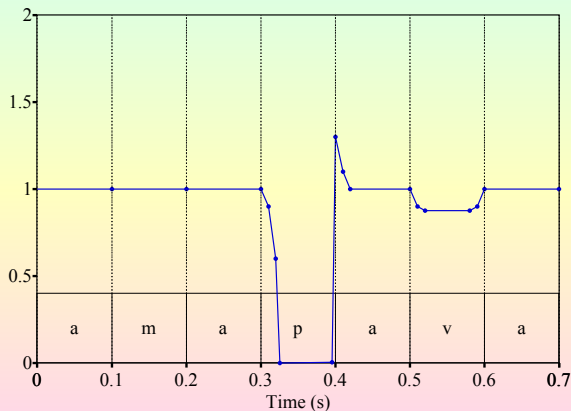


Figure: Model of micromelodic profile for /amapava/

Macromelodic profile

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

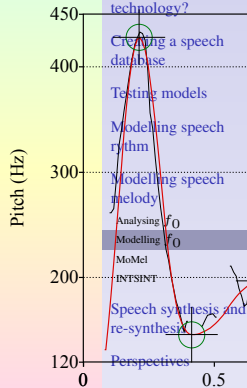
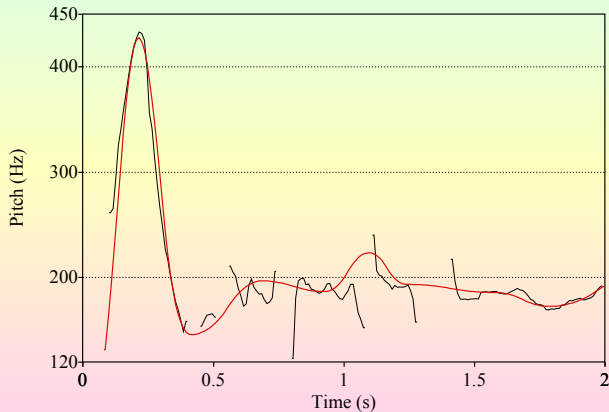


Figure: Macromelodic profile for extract from A01-01

f_0 transitions

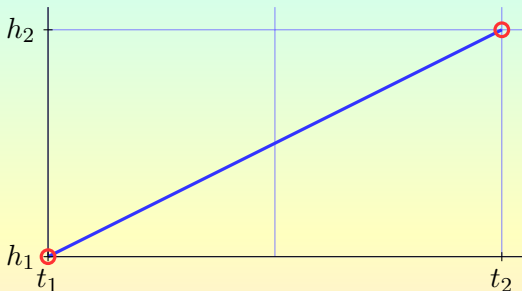


Figure: An f_0 transition from $\langle t_1, h_1 \rangle$ to $\langle t_2, h_2 \rangle$

Linear transition:

$$h_i = h_1 + \frac{(t_2 - t_i)}{(t_2 - t_1)} \cdot (h_2 - h_1)$$

Other functions

Normalising with $t_1, h_1 = 0$ and $t_2, h_2 = 1$

linear $y = x$

cosine $y = (1 - \cos(\pi x))/2$

critically damped harmonic $y = 1 - (1 + kx)e^{-kx}$

3rd degree polynomial $y = 3x^2 - 2x^3$

Gompertz $y = a^{b^x}$

logistic $y = 1/(a \cdot b^x + 1)$

hyperbolic $y = (\tanh(ax + b) + 1)/2$

etc. ...

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

First derivative of raw f_0

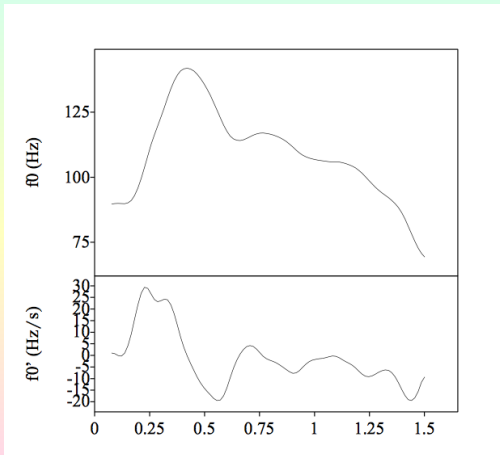


Figure: “But who stole Jane’s bicycle?” /ma’mɑ’mɑ’mɑ’mamama/

Quadratic spline function

spline function continuous function of degree n

the derivatives of which up to $n - 1$ are everywhere continuous.

cubic spline commonly used for interpolating missing values

quadratic spline smooth monotonic interpolation between points (**targets**) where derivative = 0.

linear spline piecewise linear
= derivative of quadratic spline

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

f_0 derivative

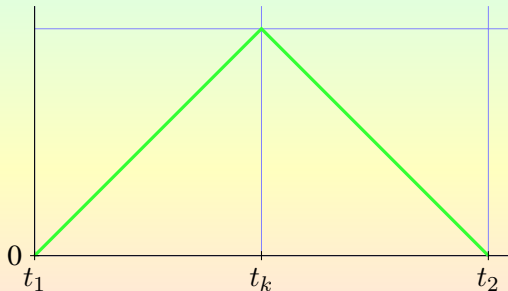


Figure: An f_0 derivative between two targets

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

f_0 transitions

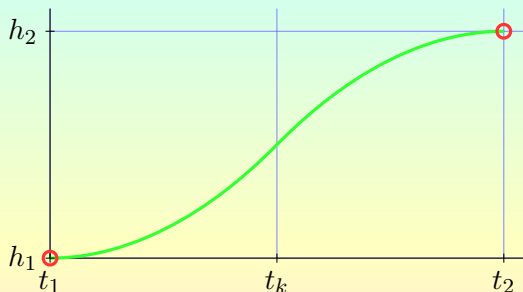


Figure: An f_0 transition from $\langle t_1, h_1 \rangle$ to $\langle t_2, h_2 \rangle$

Quadratic transition :

$$t_i \in [t_1 \dots t_k] : h_i = h_1 + \frac{(h_2 - h_1) \cdot (t_i - t_1)^2}{(t_k - t_1)(t_2 - t_1)}$$

$$t_i \in [t_k \dots t_2] : h_i = h_2 + \frac{(h_1 - h_2) \cdot (t_i - t_2)^2}{(t_k - t_2)(t_1 - t_2)}$$

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

MoMel

An algorithm for modelling melody.

Manual momel Used from 1980 on to model melody

Automatic momel Hirst & Espesser 1993

Regression variety of robust regression

Quadratic First derivative is linear

Asymmetric Microprosody is essentially a lowering of f_0

Modal generalisation of mode to function

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Mean and Mode

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

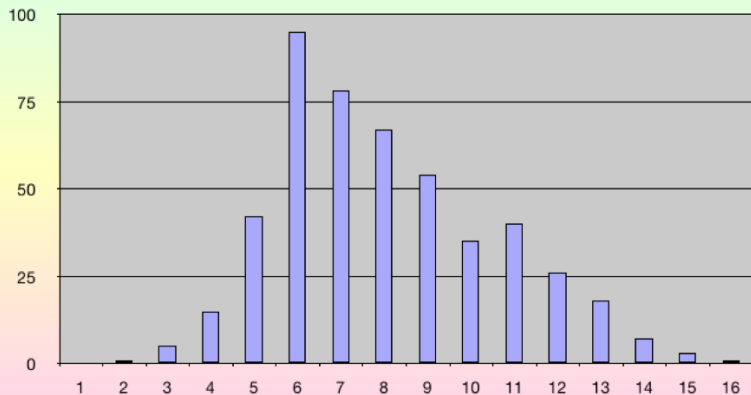


Figure: Mean and Mode of distribution

Mean and Mode

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Parameters of distribution

Mean value with minimum sum of squares of
differences from data

Mode value with maximum number of values less than
 δ from data

Generalise to function

Standard regression function with minimum sum of squares of
differences from data

Modal regression function with maximum number of values
less than δ from data

Macromelodic profile

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

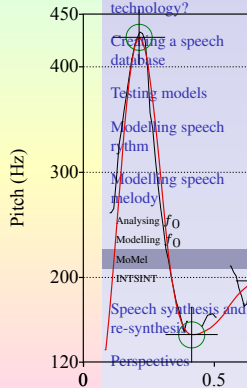
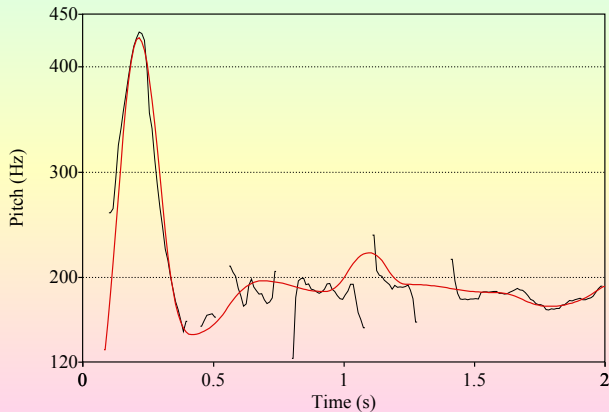


Figure: Macromelodic profile for extract from A01-01

Evaluation of Momel

Corpus	Lang.	Nombre de points			Evaluation				
		<i>auto</i>	<i>ajout.</i>	<i>suppr.</i>	<i>silence</i>	<i>bruit</i>	<i>rappel</i>	<i>précis.</i>	<i>F</i>
<i>Eurom</i>	en	8380	623	125	7,0	1,5	93,0	98,5	95,7
	fr	6547	423	130	6,2	2,0	93,8	98,0	95,9
	ge	13595	1145	506	8,0	3,7	92,0	96,3	94,1
	it	9475	337	330	3,6	3,5	96,4	96,5	96,5
	sp	8985	651	16	6,8	0,2	93,2	99,8	96,4
	toutes	46982	3179	1107	6,5	2,4	93,5	97,6	95,5
<i>Fref</i>	fr	9835	532	744	5,5	7,6	94,5	92,4	93,4

Tableau 7. Evaluation de la stylisation automatique.

Figure: Estelle Campione (2001)

Praat plugin

Old algorithm

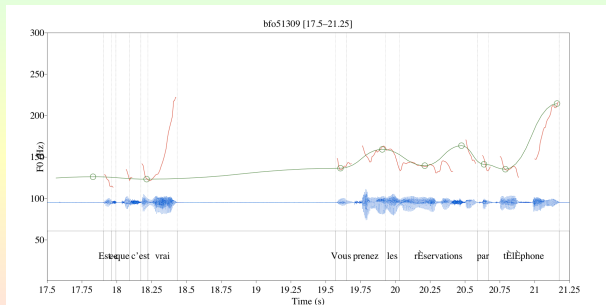


Figure: “Est-ce que c’est vrai ? Vous prenez les réservations par téléphone ?” Old version

Praat plugin

New algorithm

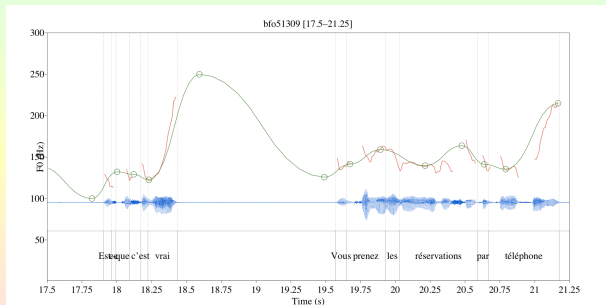


Figure: “Est-ce que c’est vrai ? Vous prenez les réservations par téléphone ?” New version

Theory neutral?

- ▶ Theory-friendly
- ▶ Phonetic representation - first step for:
 - ▶ Fujisaki model (Mixdorff)
 - ▶ ToBI
 - ▶ (Maghbouleh, Wightman & Campbell, Cho (K-ToBI))
 - ▶ INTSINT

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and re-synthesis

Perspectives

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

An INternational Transcription System for INTonation.

- ▶ Based on minimal pitch contrasts in descriptions of intonation patterns
- ▶ Used in Hirst & Di Cristo (eds) 1998 for 9 different languages
- ▶ Extension for duration and rhythm (Hirst 1999)

Basic INTSINT

Absolute tones T(op) M(id) B(ottom)

Relative tones H(igher) S(ame) L(ower)

Iterative relative tones U(pstepped) D(ownstepped)

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

INTSINT to MoMel

Speech Technology
for Human Scientists
94/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

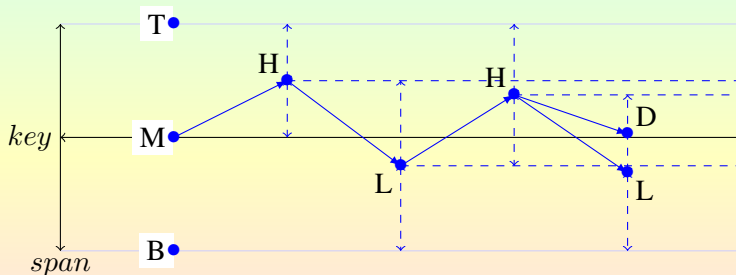


Figure: INTSINT to MoMel defined by 2 parameters *key* and *span*

Downdrift

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

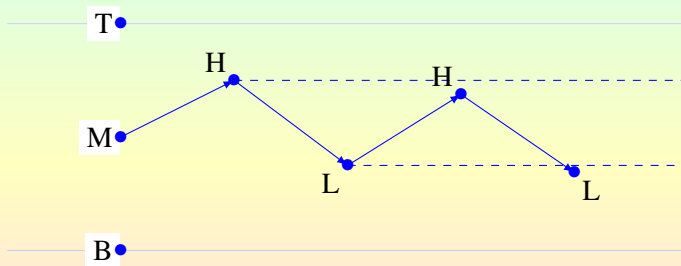


Figure: Downtdrift is an automatic by-product of the way in which the relative tones are defined

MoMel to INTSINT

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

INTSINT

Speech synthesis and
re-synthesis

Perspectives

- ▶ Perl script
- ▶ Optimal coding within target space:
 - key** mean ± 50 Hz (step: 1)
 - span** 0.5...2.5 octaves (step: 0.1)
- ▶ for each couple $\langle key, span \rangle$ find optimal coding
- ▶ retain optimal parameters and coding

INTSINT output

```
; A01_01.intsint created on Tue Aug 24 08:12:47 2010 by intsint.pl 2.11
; from A01_01.momel
; 32 values mean = 191
<parameter range=1.4>
<parameter key=235>
0.113 M 221 235
0.219 D 205 208
0.434 D 182 190
0.746 B 120 145
1.177 S 120 145
1.423 T 428 382
1.623 B 146 145
1.894 U 197 184
```

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Analysing f_0

Modelling f_0

MoMel

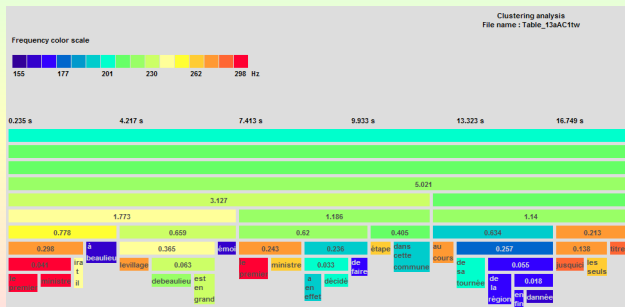
INTSINT

Speech synthesis and
re-synthesis

Perspectives

Long term characteristics

INTSINT supposes that there are no variations in *key* and *span*
In authentic speech this is obviously not true.



Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Synthesis systems

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

Mbrola Diphone synthesis

Large collection of languages available

Festival complex modular system

Praat PSOLA (Pitch Synchronous Overlap and Add)

ProZed uses Praat

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

ProZed

Perspectives

ProZed

ProZed

rhythm linear model

melody Intsint model

Speech Technology
for Human Scientists
102/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

ProZed - rhythm

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

A linear model (Hirst & Auran 2005)

t coefficient of tempo

k scalar lengthening

q quantal lengthening

$$\hat{d}_\rho = t \cdot \left\{ \sum_{i=1}^m \bar{d}_{i/p} + k \cdot q \right\}$$

Predicted vs observed

duration

Speech Technology
for Human Scientists
104/109

Daniel Hirst

Outline

Introduction: Why

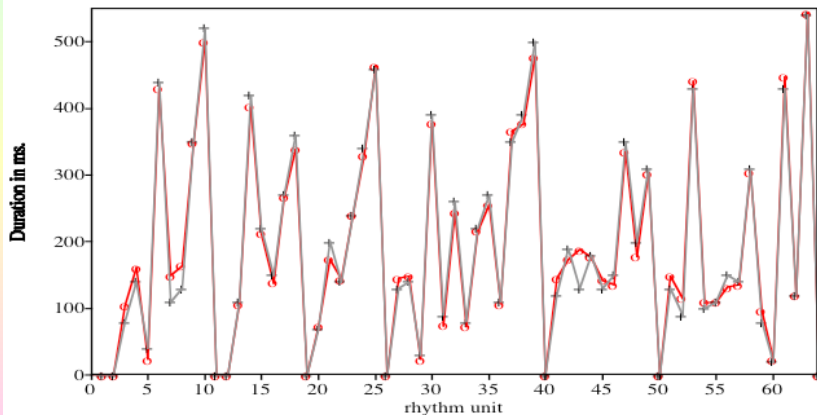


Figure: Predicted versus observed values for duration

ProZed

rhythm

Speech Technology
for Human Scientists
105/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

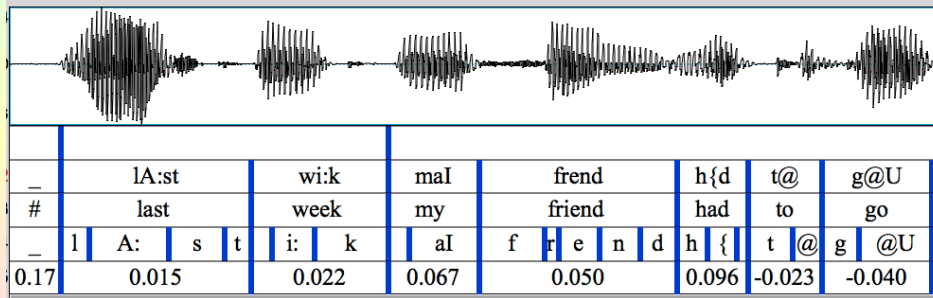


Figure: Using the ProZed plugin to model rhythm

Predicted vs observed

pitch targets

Speech Technology
for Human Scientists
106/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

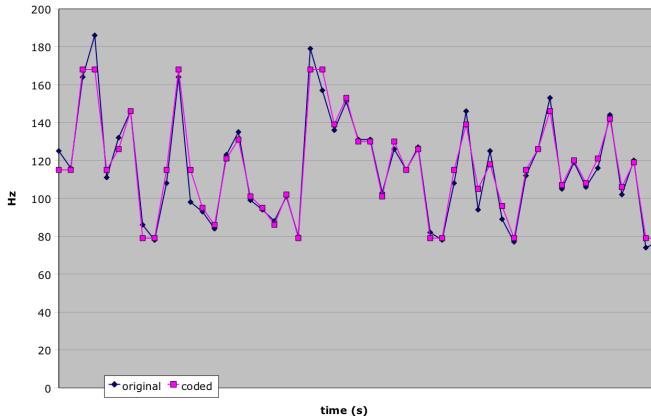


Figure: Predicted versus observed values for pitch targets

ProZed

melody

Speech Technology
for Human Scientists
107/109

Daniel Hirst

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

ProZed

Perspectives

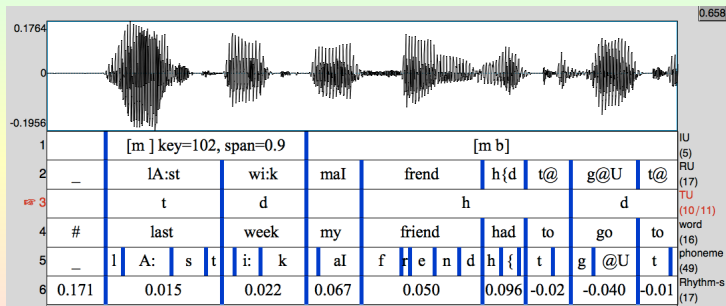


Figure: Using the ProZed plugin to model melody

Outline

Introduction: Why
do we need
technology?

Creating a speech
database

Testing models

Modelling speech
rhythm

Modelling speech
melody

Speech synthesis and
re-synthesis

Perspectives

Introduction: Why do we need technology?

Creating a speech database

Testing models

Modelling speech rhythm

Modelling speech melody

Speech synthesis and re-synthesis

Perspectives

Questions

If I don't have time to answer your questions today, then

- ▶ see me during the conference, or
- ▶ email me: daniel.hirst@lpl-aix.fr
- ▶ You can download my Praat plugins and other files from:

http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst