The Roles of a Generative Model in the Study of Tonal Features of Speech

Hiroya Fujisaki

Professor Emeritus, University of Tokyo
fujisaki@alum.mit.edu
Outline

1. Introduction — The roles of models and the Importance of physiology, physics, and mathematics in model construction
2. A quantitative model for the generation process of $F_0$ contours of Japanese
3. Physiological and physical mechanisms
4. Applications
   (a) Prosodic organization of spoken Japanese
   (b) Analysis and synthesis of $F_0$ contours of other languages
   (c) Speech recognition
   (d) Typology of tonal features of languages
   (e) Phonology of tone systems of some tone languages
5. Conclusions
# Information Conveyed by Speech

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Discrete/Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic</td>
<td>Lexical (word accent, etc.)</td>
<td>Discrete (symbolic)</td>
</tr>
<tr>
<td></td>
<td>Syntactic (phrase structure, etc.)</td>
<td>Controlled by speaker</td>
</tr>
<tr>
<td></td>
<td>Pragmatic (discourse focus, etc.)</td>
<td></td>
</tr>
<tr>
<td>Para-linguistic</td>
<td>Intentional (exhortation, etc.)</td>
<td>Discrete and Continuous</td>
</tr>
<tr>
<td></td>
<td>Attitudinal (politeness, etc.)</td>
<td>Can be controlled by speaker</td>
</tr>
<tr>
<td></td>
<td>Stylistic (fast, slow, etc.)</td>
<td></td>
</tr>
<tr>
<td>Non-linguistic</td>
<td>Physical (age, gender, etc.)</td>
<td>Discrete and/or Continuous</td>
</tr>
<tr>
<td></td>
<td>Emotional (joy, sorrow, etc.)</td>
<td>Generally cannot be controlled but can be simulated</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic</td>
<td></td>
</tr>
</tbody>
</table>
Processes of Information Manifestation by Speech (Fujisaki 1995)

Input Information
- Linguistic
  - Lexical
  - Syntactic
  - Semantic
  - Pragmatic
- Paralinguistic
  - Intentional
  - Attitudinal
  - Stylistic
- Non-linguistic
  - Physical
  - Emotional

Rules of Grammar → Message Planning
Rules of Prosody → Utterance Planning
Physiological Constraints → Motor Command Generation
Physical Constraints → Speech Sound Production

Segmental and Suprasegmental Features of Speech

© Hiroya Fujisaki
Role of Voice Fundamental Frequency ($F_0$) Contour

• In many languages, the $F_0$ contour is used to express **tone**, **accent**, and **intonation**, and plays a major role in conveying linguistic information on the prosody (i.e., the structural organization of various linguistic units into a coherent utterance or a coherent group of utterances).

• It can convey also **para-linguistic** information concerning speaker’s intention and attitude, as well as **non-linguistic** information concerning speaker’s physical and mental states (such as age, emotion, etc.)
### Three Approaches for Description/Representation of $F_0$ Contour Characteristics

<table>
<thead>
<tr>
<th>Example</th>
<th>Outcome</th>
<th>Method</th>
<th>Coding/Decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeling</td>
<td>Discrete Labels</td>
<td>Subjective Qualitative</td>
<td>Irreversible</td>
</tr>
<tr>
<td>Stylization</td>
<td>Piece-wise Linear Approx.</td>
<td>Subjective Quantitative</td>
<td>Irreversible</td>
</tr>
<tr>
<td>Modeling</td>
<td>Timing and Magnitude of Commands</td>
<td>Objective Quantitative</td>
<td>Reversible</td>
</tr>
</tbody>
</table>

© Hiroya Fujisaki

TAL 2006 (2006-04-27)
1. Essential characteristics of the structure of a mechanism or the function of a process (in speech production and speech perception).

2. Formal description (representation) of a set of rules and constraints, or the elements and structures (in a theory of generative grammar).
Requirements for a Good Model

1. *Objective*: Its parameters should be obtainable by objective methods

2. *Quantitative*: It should capture both *discrete* and *continuous* features of intonation

3. *Generative*: It should be capable of reproducing/generating the entire $F_0$ contour from a set of parameters

4. *Explicable*: In terms of information expressed by the message
   - In terms of the underlying mechanism
Role of Generative Models in Research
(Fujisaki 1995)

1. Abduction
   Finding the model
   
2. Deduction/Synthesis
   Testing the model
   
3. Induction/Recognition
   Inferring the underlying events

© Hiroya Fujisaki
Lexical Word Accent in Japanese

1. The subjective pitch of each *mora* is either H or L. The pattern of subjective pitch of a word constitutes a “word accent type”, and each dialect is characterized by its own system of word accent types.

2. In common Japanese (based on the *Tokyo* dialect), an upward transition (L to H) can occur only at the end of the initial *mora*, and no more than one downward transition (H to L) is allowed in a word. Thus there are \( n+1 \) accent types for \( n \)-*mora* words.

3. The type with a downward transition at the end of the \( i \) th *mora* is denoted by \((n, i)\).
# Subjective Pitch Patterns of Word Accent Types in Common Japanese

<table>
<thead>
<tr>
<th>Word accent type</th>
<th>Number of <em>morae</em> in a word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>i</td>
</tr>
<tr>
<td>1</td>
<td>o</td>
</tr>
<tr>
<td>2</td>
<td>i – e</td>
</tr>
<tr>
<td>3</td>
<td>u – re – i</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
**F₀ Contours of 5-mora Words of Common Japanese**

![Graph of F₀ contours for 5-mora words](image)

- Type (5,0) a-wa-re-mo-no
- Type (5,1) a-me-a-ra-re
- Type (5,2) a-o-i-u-o
- Type (5,2) a-ma-ga-e-ru
- Type (5,4) mi-a-ya-ma-ru
- Type (5,5) o-sho-ga-tsu

Average logarithmic F₀ contours for all the accent types of 5-mora words.
A Model for Generation of $F_0$ Contour of a Spoken Word (Fujisaki 1969)
Example of a measured $F_0$ contour of the declarative sentence:

"Aoi aoinoewa yamanouenoieni aru."

(The picture of the blue hollyhock is in a house on top of the hill.)
A model for the process of generating the $F_0$ contour of a spoken sentence (Fujisaki et al. 1982)

$$\text{log}_e F_0(t) = \text{log}_e F_b + \sum_{i=1}^{I} A_{p_i} G_p(t-T_{0i}) + \sum_{j=1}^{J} A_{a_j} \{G_a(t-T_{1j}) - G_a(t-T_{2j})\}$$  (1)

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases}$$  (2)

$$G_a(t) = \begin{cases} \text{Min}[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\ 0, & t < 0. \end{cases}$$  (3)
**Phrase and Accent Components with Typical Values of ћ, ќ and Ѝ**

Parameter values for the phrase component: ћ = 3.0/s, the accent components: ќ = 20.0/s, Ѝ = 0.9.
Use of a Generative Model in Synthesis and Analysis

SYNTHESIS

UNDERLYING COMMANDS

MODEL

MANIFESTATIONS (F₀ CONTOUR)

IN

OUT

ANALYSIS

COMMANDS

 snoke

MEASURED F₀ CONTOUR

COMPARE

SYNTHESIZED F₀ CONTOUR

MODEL

GENERATE COMMANDS

INVERSE PROBLEM

IN

OUT

ANALYSIS-BY-SYNTHESIS
Analysis-by-Synthesis of an $F_0$ Contour of Common Japanese
Rationale for the Existence of the Second Phrase Command

Analysis-by-Synthesis of the $F_0$ contour of "aoi aoinewa yamanouenoi ni aru".
Analysis-by-Synthesis of $F_0$ Contours of 10 Declarative Sentences of Japanese

© Hiroya Fujisaki
Roles of Physiology, Physics and Mathematics in the Model
Question (1)

Why does the model’s formulation work well if we use the logarithmic scale rather than the linear scale for $F_0$?
Structure and Function of Larynx

- PHARYNX
- LARYNX
- TRACHEA
- VOCAL CORDS (VOCALIS MUSCLES)
  - THYROID CARTILAGE
  - ARYTENOID CARTILAGES
  - VOCAL CORDS
  - CRICO-THYROID MUSCLE
  - CRICOID CARTILAGE
  - CRICO-THYROID JOINT
Stiffness as function of tension at rest (---) and during isometric tetanic contraction initiated at different original lengths. In the top curve contraction is initiated at a length below 100 (equilibrium length = 100).

Ordinate: stiffness in arbitrary units.
Abscissa: tension in arbitrary units.

Physical Properties of Skeletal Muscles (2)

\[ \frac{dT}{dx} = b(T + a) \]

\[ T = a(e^{bx} - 1) \]
Stress-strain relationship in a skeletal muscle (i.e., vocalis)

\[ \frac{dT}{dl} = b \left( T + a \right), \quad (1) \]

where \( T \): tension, \( l \): length of vocalis, \( a \): stiffness at \( T = 0 \).

By integration, we obtain

\[ T = \left( T_0 + \frac{a}{b} \right) \exp \left\{ b(l - l_0) \right\} - \frac{a}{b}, \quad (2) \]

where \( T_0 \): static tension, \( l_0 \): vocalis length at \( T = T_0 \).

When \( T_0 \gg \frac{a}{b} \),

\[ T \cong T_0 \exp \left( b x \right), \quad (3) \]

where \( x = l - l_0 \).
From Vocal Cord Elongation to Fundamental Frequency

Frequency of vibration of an elastic membrane is given by

\[ F_0 = c_0 \left\{ \frac{T}{\sigma} \right\}^{1/2}, \quad \text{where } \sigma : \text{density/unit area.} \quad (4) \]

From Eqs. (3) and (4) we obtain

\[ \log_e F_0 = \log_e \left[ c_0 \left\{ \frac{T_0}{\sigma} \right\}^{1/2} \right] + \left( \frac{b}{2} \right) x. \quad (5) \]

When \( x \) is time-varying, i.e., \( x = x(t) \),

\[ \log_e F_0(t) = \log_e F_b + \left( \frac{b}{2} \right) x(t), \quad (6) \]

where \( F_b = c_0 \left\{ \frac{T_0}{\sigma} \right\}^{1/2} \).

Thus an \( F_0 \) contour, when plotted in the logarithmic scale as a function of time, can be expressed as the sum of a constant (baseline) term and a time-varying term, the latter being proportional to the elongation of the vocal cord.
Question (2)

Why does the accent component of log $F_0 (t)$ have a shape similar to that of the step response of a second-order linear system (i.e., a system consisting of a mass, a viscous loss, and a spring)?
Thyroid and Cricoid Cartilages as a Two-Mass, Two-Spring System

VOC: Vocalis M.
CT: Cricothyroid M.
T: Thyroid
C: Cricoid
A: Arytenoid

l: Length of vocalis
x: Elongation of vocalis
θ: Angular displacement of thyroid
Rotation of Thyroid Around the Crico-Thyroid Joint

\[ I \frac{d^2 \theta}{dt^2} + R \frac{d\theta}{dt} + K \theta = \tau(t), \]

where \( \tau(t) \): Torque generated by contraction of CT

thus \( \theta(t) = C_3 G_a(t) \).

• For small \( \theta \),
  \[ x(t) = C_4 \theta(t) = C_5 G_a(t) \]

• Hence
  \[ \log_e F_0(t) = C_6 G_a(t) + C \]
The rate of $F_0$ change is determined, not by the speed of contraction or relaxation of the muscle, but by the mechanical properties of the laryngeal structure.

The rate of change varies with the amplitude of the command, but the time constant remains the same. (Fujisaki, 1981)
**Global and Local Components of an $F_0$ Contour**

Almost all the studies on $F_0$ contours assume, either explicitly or implicitly, the presence of at least two separate phenomena which differ in their range of influence on $F_0$.

<table>
<thead>
<tr>
<th>Global</th>
<th>phrase level phenomena</th>
<th>phrase components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>word/syllable level phenomena</td>
<td>accent components/ tone components</td>
</tr>
</tbody>
</table>

© Hiroya Fujisaki
Question (3)

Why does the phrase component of log $F_0(t)$ have a shape similar to that of the impulse response of another second-order linear system (i.e., a system consisting of a mass, a viscous loss, and a spring)?

And why are the phrase components and accent components additive in log $F_0(t)$?
**Motion of Thyroid with Two Degrees of Freedom**

- **Rotation of thyroid by** *pars recta* of the cricothyroid muscle
- **Translation of thyroid by** *pars obliqua* of the cricothyroid muscle

---

© Hiroya Fujisaki
Rotation and Translation of Thyroid

Rotation

\[ M r^2 \frac{d^2 \theta}{dt^2} + R \frac{d\theta}{dt} + K \theta = \tau(t) \]

\( \tau(t) \) : Torque generated by contraction of CT pars recta

Translation

\[ M \frac{d^2 x}{dt^2} + R' \frac{dx}{dt} + K' x = f(t) \]

\( f(t) \) : Force generated by contraction of CT pars obliqua

© Hiroya Fujisaki
Additivity of Phrase and Accent Components

\[ F_0 \text{ contour} \]
\[ \log F_0(t) \square \{ x_1(t) + x_2(t) + c \} \]
The term “Supersposition” applies only if one uses the logarithmic (i.e., semitone) scale for $F_0$. 

The diagram shows the additivity of components in the log$F_0$ domain. 

Phrasing: 
- Activity of CT *pars obliqua* 
- Translation of thyroid 
- Small change ($x_1$) in vocal cord length 
- Resultant change ($x_1 + x_2$) in vocal cord length 
- Change in log $F_0$ proportionate to ($x_1 + x_2$) 

Accentuation: 
- Activity of CT *pars recta* 
- Rotation of thyroid 
- Small change ($x_2$) in vocal cord length
From Linguistic Information to Acoustic-Phonetic Manifestations of Word Accent and Sentence Intonation

<table>
<thead>
<tr>
<th>Observed Phenomena</th>
<th>Related Academic Disciplines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word accent and sentence intonation</td>
<td>linguistics and phonetics</td>
</tr>
<tr>
<td>(neuromotor commands)</td>
<td></td>
</tr>
<tr>
<td>Cricothyroid muscle activity</td>
<td>physiology</td>
</tr>
<tr>
<td>(contractile force)</td>
<td></td>
</tr>
<tr>
<td>Thyroid cartilage movements</td>
<td>physics</td>
</tr>
<tr>
<td>(dynamics of rigid bodies)</td>
<td></td>
</tr>
<tr>
<td>Changes in vocal cord length</td>
<td>mathematics</td>
</tr>
<tr>
<td>(geometry)</td>
<td></td>
</tr>
<tr>
<td>Changes in vocal cord tension</td>
<td>physiology / physics</td>
</tr>
<tr>
<td>(elasticity)</td>
<td></td>
</tr>
<tr>
<td>Changes in fundamental frequency</td>
<td>physics / acoustics</td>
</tr>
<tr>
<td>(vibration of elastic membrane)</td>
<td></td>
</tr>
</tbody>
</table>

© Hiroya Fujisaki
Role of the Model as a Tool for Investigating the Units and Structures of Prosody
Use of Phrase Components for **Disambiguation of Ambiguity due to Constitutional Homonymity** (Fujisaki, 1981)

(a) Left-branching phrase structure (one phrase component)

(b) Center-embedded phrase structure (two phrase components)

Analysis-by-Synthesis of $F_0$ Contours of Sentences with Constitutional Homonymity
‘Declination’ caused by phrase components

© Hiroya Fujisaki
1. A **prosodic word** is defined as a part or a whole of an utterance that forms an accent type in a specific dialect of spoken Japanese. (cf. Selkirk 1979)

2. A **prosodic phrase** is defined as the interval between two successive phrase commands uninterrupted by a pause.

3. A **prosodic clause** is defined as a successive set of prosodic phrases delimited by utterance-medial/final pauses.

4. A **prosodic sentence** is defined as the utterance delimited by utterance-final pauses (except for the initial utterance of a discourse segment).
### Hierarchy of Prosodic Units of Spoken Japanese (Fujisaki 1988)

<table>
<thead>
<tr>
<th>MANIFESTATIONS</th>
<th>PROSODIC UNITS</th>
<th>SYNTACTIC UNITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN ACCENT COMPONENT</td>
<td>PROSODIC WORD</td>
<td>‘BUNSETSU’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(a content word followed by (n \geq 0) function words)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>sandhi</em></td>
</tr>
<tr>
<td>A PHRASE COMPONENT</td>
<td>PROSODIC PHRASE</td>
<td><em>ICRLB</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Immediate Constituents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with Recursively</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left-branching Structure)</td>
</tr>
<tr>
<td>PHRASE COMPONENT(S) DELIMITED BY</td>
<td>PROSODIC CLAUSE</td>
<td>CLAUSE</td>
</tr>
<tr>
<td>UTTERANCE-MEDIAL PAUSE(S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHRASE COMPONENT(S) DELIMITED BY</td>
<td>PROSODIC SENTENCE</td>
<td>SENTENCE</td>
</tr>
<tr>
<td>UTTERANCE-FINAL PAUSES</td>
<td>(UTTERANCE)</td>
<td></td>
</tr>
</tbody>
</table>
Syntactic Structure vs. Prosodic Structure

**SYNTACTIC STRUCTURE**

その上 3月 20日には、東京の 地下鉄の 車内で 毒ガスの サリンが 撒かれるという、前代未聞の 怪事件も 起き、

**PROSODIC STRUCTURE**

PW: Prosodic Word
PP: Prosodic Phrase
PC: Prosodic Clause

© Hiroya Fujisaki

TAL 2006 (2006-04-27)
Prosodic Structure of Japanese Text Reading

written text

spoken discourse

paragraph

DS 1 DS 2 DS 3

DS: discourse segment

P1 (pause between paragraphs)

sentence

SS 1 SS 2 SS 3

SS: spoken sentence

P2 (pause between sentences)

clause

PC 1 PC 2 PC 3

PC: prosodic clause

P3 (pause within a sentence)

phrase

PP1 PP2 PP3

PP: prosodic phrase

word

PW1 PW2 PW3

PW: prosodic word

© Hiroya Fujisaki

TAL 2006 (2006-04-27) 45
Application to Languages Other than Japanese
Objective: To find out
- To what extent the original model for Japanese will apply to other languages
- What are the modifications, if any, necessary for each language
- Both universal and language-specific characteristics of $F_0$ control

Languages investigated up to present:
Chinese (Mandarin, Cantonese and Shanghainese), English, Estonian, German, Greek, Hindi, Korean, Polish, Portuguese, Russian, Spanish, Swedish, Thai, Vietnamese

Speech Material:
- Readings of sentences, paragraphs, and short stories
- Recorded by native speakers of each language
Application of the Model to Languages other than Japanese

English

German

Greek

Korean

Polish

Spanish

© Hiroya Fujisaki
The Four Tones of Standard Chinese

<table>
<thead>
<tr>
<th>Tone</th>
<th>Name</th>
<th>Description</th>
<th>Symbol</th>
<th>Semi–Quantitative Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>阴平声(第一声)</td>
<td>高平</td>
<td>Ɩ</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>陽平声(第二声)</td>
<td>高降</td>
<td>Ɩ</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>上声(第三声)</td>
<td>降昇</td>
<td>√</td>
<td>214</td>
</tr>
<tr>
<td>4</td>
<td>去声(第四声)</td>
<td>仝降</td>
<td>√</td>
<td>51</td>
</tr>
</tbody>
</table>

The syllable “yi” uttered with four different $F_0$ contours, and its four different meanings.
Tone Contours and the Underlying Tone Commands

\[ F_0(t) [\text{Hz}] \]

(Tone 1) \hspace{2cm} (Tone 2) \hspace{2cm} (Tone 3) \hspace{2cm} (Tone 4)

Tone Command

Polarity of Commands

© Hiroya Fujisaki
Model for $F_0$ Contours of Standard Chinese (Fujisaki et al. 1987)

- Correspond to Chao’s Large Waves
- Corresponds to Wu’s Change-Key
- Correspond to Chao’s Small Ripples

© Hiroya Fujisaki
Model with Positive and Negative Tone Commands

\[
\log_e F_0(t) = \log_e F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) \\
+ \sum_{j=1}^{J} A_{tj} \{ G_t(t - T_{1j}) - G_t(t - T_{2j}) \}, \tag{8}
\]

where

\[
G_p(t) = \begin{cases} 
\alpha^2 t \exp(-\alpha t), & t \geq 0, \\
0, & t < 0,
\end{cases} \tag{9}
\]

\[
G_t(t) = \begin{cases} 
\min[1 - (1 + \beta_1 t) \exp(-\beta_1 t), \gamma_1], & t \geq 0, \\
0, & t < 0,
\end{cases} \tag{10}
\]

(for positive tone commands)

\[
G_t(t) = \begin{cases} 
\min[1 - (1 + \beta_2 t) \exp(-\beta_2 t), \gamma_2], & t \geq 0, \\
0, & t < 0,
\end{cases} \tag{10}
\]

(for negative tone commands).
### Parameters and Constants of the Model

- $F_b$: baseline value of fundamental frequency,
- $I$: number of phrase commands,
- $J$: number of tone commands,
- $A_{pi}$: magnitude of the $i$th phrase command,
- $A_{tj}$: amplitude of the $j$th tone command,
- $T_{0i}$: timing of the $i$th phrase command,
- $T_{1j}$: onset of the $j$th tone command,
- $T_{2j}$: end of the $j$th tone command,
- $\alpha$: natural angular frequency of the phrase control mechanism,
- $\beta_1$: natural angular frequency of the tone control mechanism to positive tone commands,
- $\beta_2$: natural angular frequency of the tone control mechanism to negative tone commands,
- $\gamma_1$: relative ceiling level of positive tone components,
- $\gamma_2$: relative ceiling level of negative tone components.
Chinese (Mandarin)

Mu4 hei1 lan3 bu2 shi4 zi3 hua1 ni2 buo2 hui4 kui4 dian4 wan4 tong3.
Question (4)

What is the mechanism responsible for the production of negative components of log $F_0(t)$ in tones of Chinese and Thai?
Midsagittal Section of the Vocal Tract and Surrounding Structures
Extrinsic Muscles of the Tongue as Viewed in Lateral Dissection

- Mastoid process
- Styloid process
- Posterior belly of digastric muscle
- Palatoglossus muscle
- Styloglossus muscle
- Hyoglossus muscle
- Genioglossus muscle
- Geniohyoid muscle
- Mylohyoid muscle
- Anterior belly of digastric muscle
- Hyoid bone
Infrahyoid Strap Muscles
Infrahyoid Muscle Activities During Production of Thai Tones (Speaker 2) (Erickson ’76)
Mechanism of $F_0$ Lowering by Activities of TH and SH

TEMPORAL BONE  MANDIBLE

LIGAMENT  HYOID

TRITICIAL CARTILAGE

LIGAMENT

THYROID  SH

VOC

CRICOID  CT

STERNUM

TH: thyrohyoid m.

SH: sternohyoid m.

CT: cricothyroid m.
Presence of Negative Accent Commands in (Some) Non-tone Languages

- Swedish
- Hindi
- Portuguese

but also

- British English
- Russian
Application of the Model to Languages other than Japanese

**Swedish**

![Swedish phonetic analysis](image)

**Hindi**

![Hindi phonetic analysis](image)

**Portuguese**

![Portuguese phonetic analysis](image)
You always try to do **everything** at the last minute.
Russian
<table>
<thead>
<tr>
<th>Group</th>
<th>Tone/Accent Command Polarity</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>positive only</td>
<td>English***, Estonian, German, Greek, Japanese, Korean, Polish, Spanish</td>
</tr>
<tr>
<td>2</td>
<td>positive, zero and negative</td>
<td>Cantonese**, Mandarin**, Thai**, Hindi, Portuguese, Russian, Swedish</td>
</tr>
</tbody>
</table>

* positive/zero/negative  ** Tone languages  *** Certain speakers of English (both American and British) occasionally use negative accent commands, especially in order to express paralinguistic information.
(a) Mandarin  
(b) Shanghainese  
(c) Thai  
(d) Vietnamese  
(e) Lanqi

○: indicates the overlapping of two tones types, e.g., entering and non-entering tones.
1. Analysis-Resynthesis with Original $F_0$ Contour
2. Analysis-Resynthesis with Model-Generated $F_0$ Contour
   (Best approximation to the original $F_0$ contour)
3. Analysis-Resynthesis with Model-Generated $F_0$ Contour
   (Quantization of accent/tone command amplitude)

Languages with
(a) Only positive accent commands:
   Japanese, English, German, Greek, Korean, Spanish
(b) Both positive and negative tone/accent commands:
   Chinese, Swedish
   British English (some utterances)
Demonstration of F₀ Contour Synthesis (2)


2. English: It’s strange that I slept so long since I wasn’t feeling tired.

3. German: Sie haben den Wagen geliehen und sind tatsächlich gefahren.

4. Greek: Είναι το σύμβολο όχι μόνο της Αθήνας, αλλά και της Ελλάδας.


7. Chinese: Mu ni hei buo lan hui bu kui shi dian zi wan hua tong.

Demonstration of $F_0$ Contour Synthesis (3)

Examples of utterances with negative accent commands in British English:

1. *Where are you going?*

2. *Have you made up your mind?*

3. *You always try to do everything at the last minute.*

© Hiroya Fujisaki
Conclusions (1)

1. The contour of voice fundamental frequency, when expressed in the logarithmic scale, generally consists of two types of components:
   - global components for phrase-level intonation,
   - local components for syllable tone/word accent.

2. The origin of these components can be explained by referring to the physiological and physical mechanisms for controlling the tension of the vocalis muscle.

3. The functions of the mechanisms can be expressed quite accurately by the command-response model.
Conclusions (2)

4. The command-response allows one to clarify the units and structures of prosody.

5. The tonal characteristics of various languages can be typologically classified according to the polarity of commands for the local components of their $F_0$ contours.

6. The command-response model can generate $F_0$ contours of speech, not only of Japanese, but also of a number of other languages, with a very high naturalness, and is especially useful in speech synthesis.