Physiology, Physics and Mathematics of $F_0$ contour Generation
with Application to Phonological Representation of Tone Languages

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 modeling

2. A quantitative model for the generation process of \( F_0 \) contours

3. Physiological and physical mechanisms, and mathematical representations of their essential characteristics

4. Applications
   (a) Disclosing units and structures of prosody
   (b) Analysis and synthesis of \( F_0 \) contours of many languages
   (c) Typology of tonal features of languages
   (d) Phonology of tone systems of some tone languages

5. Conclusions
### Information Conveyed by Speech (Fujisaki 1994)

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Discrete/Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic</strong></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><strong>Para-linguistic</strong></td>
<td>Intentional (exhortation, etc.)</td>
<td>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><strong>Non-linguistic</strong></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)
What is Prosody? --- The Author’s Definition
(Fujisaki 1995)

Prosody is defined as the systematic organization of individual linguistic units into an utterance, or a coherent group of utterances, in the process of speech production.

Its realization involves both segmental and suprasegmental features of speech, and is influenced, not only by linguistic information, but also by para-linguistic and non-linguistic information.
Role of Voice Fundamental Frequency ($F_0$) Contour

- In many languages, the pattern of temporal changes in $F_0$ (henceforth 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 to the 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>ToBI</td>
<td>Discrete Labels</td>
<td>Subjective Qualitative</td>
</tr>
<tr>
<td>Stylization</td>
<td>’t Hart</td>
<td>Piece-wise Linear Approx.</td>
<td>Subjective Quantitative</td>
</tr>
<tr>
<td>Modeling</td>
<td>Fujisaki</td>
<td>Timing and Magnitude of Commands</td>
<td>Objective Quantitative</td>
</tr>
</tbody>
</table>
What is a Model?

1. Essential characteristics of the structure of a mechanism or the function of a process (Fujisaki 1969, both in speech production and in speech perception).

2. Formal description/representation of a set of rules and constraints, or the elements and structures in a theory of generative grammar (used by Chomsky and his school since early 70s).
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. **Explanatory Power:**
   - 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

   OBSERVED PHENOMENA → HYPOTHESIS BUILDING → MODEL
   UNDERLYING EVENTS

2. Deduction/Synthesis
   Testing the model

   UNDERLYING EVENTS → MODEL → EXPECTED / OBSERVED PHENOMENA

3. Induction/Recognition
   Inferring the underlying events

   OBSERVED PHENOMENA → ANALYSIS BY SYNTHESIS → UNDERLYING EVENTS
   MODEL
Early History of Analysis-by-Synthesis in Speech Processing

1. Halle and Stevens (1959)
   Proposal as a Paradigm for Speech Processing at Various Levels
   (Symposium on Speech Communication and Processing held at Air Force Cambridge Research Center, July 1959)

2. Stevens and Halle (1960)
   Formal Presentation of the Idea
   (Trans. IEEE Prof. Group on Audio and Electroacoustics, 1960)

   First application of A-b-S to Formant Extraction (J.A.S.A., 1961)

4. Öhman and Lindqvist (1965)
   First application of A-b-S to $F_0$ contour Analysis (STL/QPSL, 1965)

5. Fujisaki and Nagashima (1969)
   Use of Fujisaki Model in $F_0$ contour parameter Extraction by A-b-S
Measured $F_0$ Contour of a Japanese Sentence (Fujisaki et al. 1979)

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)

\[
\log_e F_0(t) = \log_e F_b + \sum_{i=1}^{I} A_{pi} \cdot G_p(t-T_{0i}) + \sum_{j=1}^{J} A_{aj} \cdot \{G_a(t-T_{1j})-G_a(t-T_{2j})\}
\]

(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 $\alpha$, $\beta$ and $\gamma$

Parameter values for the phrase component: $\alpha = 3.0/\text{s}$,
the accent components: $\beta = 20.0/\text{s}$, $\gamma = 0.9$. 
Analysis-by-Synthesis of an $F_0$ Contour of Common Japanese
Analysis-by-Synthesis of F₀ Contours of 10 Declarative Sentences of Japanese

© Hiroya Fujisaki
‘Declination’ caused by phrase components

© 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

© Hiroya Fujisaki
The stress-strain relationship of skeletal muscles including the human vocalis muscle has been widely studied [e.g., Buchthal & Kaiser 1944, Sandow, 1958].

The next figure shows the earliest published data on the relationship between tension and stiffness of a skeletal muscle by Buchthal and Kaiser published in *Acta Physiol. Scand.*
Physical Properties of Skeletal Muscles (1)

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)

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

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

By integration, we obtain

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

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

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

\[ T \approx T_0 \exp (bx). \quad (3) \]

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

Frequency of vibration of an elastic membrane is given by

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

(4)

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

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

(5)

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

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

(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 with Vocalis and Cricothyroid Muscles Forming 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

- Equation of Motion (Rotation)
  \[ I \frac{d^2 \theta}{dt^2} + R \frac{d\theta}{dt} + K \theta = \tau(t), \]
  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
  \[ \log_e F_0(t) = C_6 G_a(t) + C \]

where \( G_a(t) = \begin{cases} 
  \min[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\
  0, & t < 0.
\end{cases} \)
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)
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 \( \log F_0(t) \) ?
The role of the cricothyroid (CT) muscle

Analysis of the laryngeal structure suggests that the movement of the thyroid cartilage has two degrees of freedom [e.g., Zemlin 1968, Fink & Demarest 1978].

One is rotation around the cricothyroid joint due to the activities of the pars recta of the cricothyroid muscle (henceforth CT) and the other is horizontal translation due to the activities of pars obliqua of CT.
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
**Rotation and Translation of Thyroid**

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

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

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

\( f(t) \): Force generated by contraction of CT pars obliqua
Additivity of Phrase and Accent Components

\[ x_1(t) = C_5 \, G_a(t) \]

where

\[
G_a(t) = \begin{cases} 
\min[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\
0, & t < 0.
\end{cases}
\]

\[ x_2(t) = C_7 \, G_p(t) \]

where

\[
G_p(t) = \begin{cases} 
\alpha^2 t \exp(-\alpha t), & t \geq 0, \\
0, & t < 0.
\end{cases}
\]
Additivity of Components in $\log F_0$ Domain

The term “Superposition” applies only if one uses the logarithmic (i.e., semitone) scale for $F_0$. 

© Hiroya Fujisaki
What about tone languages?
Instead of the accent commands for non-tone languages, introduce the following **tone commands** for each syllable.

![Graph showing tone contours and commands](image-url)

- **Polarity of Commands**
  - Tone 1: +
  - Tone 2: - +
  - Tone 3: -
  - Tone 4: + -
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
TAL2012, Nanjing (2012-05-27)
Model with Positive and Negative Tone Commands

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

(8)

where

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

(9)

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

(for positive tone commands),

(10)

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

(for negative tone commands).
Mandarin

Mu4 hei1 lan3 bu2 shi4 zi3 hua1 ni2 buo2 hui4 kui4 dian4 wan4 tong3.
Cantonese

WAVEFORM

FUNDAMENTAL FREQUENCY

hun3  gin3  maa2  faai3  gong2  ceot7  lai4

PHRASE COMMANDS

TONE COMMANDS

TIME [s]

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 42
Question (4)

What is the mechanism responsible for the production of negative tone components of \( \log F_0 (t) \) in tones of Mandarin, Cantonese, and many other tone languages?
Midsagittal Section of the Vocal Tract and Surrounding Structures
Extrinsic Muscles of the Tongue as Viewed in Lateral Dissection
Infrahyoid Strap Muscles

- Hyoid bone
- Thyrohyoid muscle
- Cricoid cartilage
- Thyroid cartilage
- Sternohyoid muscle
- Sternothyroid muscle
- Sternum
Infrahyoid Muscle Activities During Production of Thai Tones (Erickson ’76)

Note the simultaneous activities of SH and TH. Activity of TH is much higher than that of SH.
Mechanism of $F_0$ Lowering by Activities of TH and SH

TEMPORAL BONE    MANDIBLE

LIGAMENT

HYOID

TRITICIAL CARTILAGE

LIGAMENT

TH

TH: thyrohyoid m.

SH

SH: sternohyoid m.

VOC

THYROID

CT

CRICOID

STERNUM

CT: cricothyroid m.
Physiology and Physics of $F_0$ Control
— Summary —

\[
\log_e F_0(t) = \log_e [c_0 \{T/\sigma\}^{1/2}] = \log_e F_b + \left(\frac{b}{2}\right) x(t)
\]
Chain of causality from linguistic information to acoustic-phonetic manifestations of tone/accent and 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>
Role of the Model as a Tool for Investigating the Units and Structures of Prosody
Definition of Prosodic Units of Spoken Japanese
(Fujisaki 1988)

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\geq0) function words)</td>
</tr>
<tr>
<td>A PHRASE COMPONENT</td>
<td>PROSODIC PHRASE</td>
<td>ICRLB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Immediate Constituents with Recursively Left-branching Structure)</td>
</tr>
<tr>
<td>PHRASE COMPONENT(S) DELIMITED BY UTTERANCE-MEDIAL PAUSE(S)</td>
<td>PROSODIC CLAUSE</td>
<td>CLAUSE</td>
</tr>
<tr>
<td>PHRASE COMPONENT(S) DELIMITED BY UTTERANCE-FINAL PAUSES</td>
<td>PROSODIC SENTENCE (UTTERANCE)</td>
<td>SENTENCE</td>
</tr>
</tbody>
</table>
An Example of Disparity Between the Syntactic Structure and the Prosodic Structure

SYNTACTIC STRUCTURE

PROSODIC STRUCTURE

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

© Hiroya Fujisaki
Examples of conflicting requirements from syntax and pragmatics (Fujisaki 1983)

(a) A case where focus on the second word overrides the syntactic boundary and causes accent *sandhi*.

```
aomorino       aneno  amaguo
  aomorino aneno    amaguo
```

(b) A case where the syntax disrupts the realization of focus on the first word.

```
aneno       aono  amaguo
  aneno aono    amaguo
```
Prosodic Structure of Japanese Text Reading
(Fujisaki 1999, Fujisaki 2000)

- **Written Language**
  - text
  - paragraph
  - sentence
  - clause
  - phrase
  - word

- **Spoken Language**
  - discourse
  - DS: discourse segment
  - PS: prosodic sentence
  - PC: prosodic clause
  - PP: prosodic phrase
  - PW: prosodic word

DS: discourse segment
P1: pause between paragraphs
P2: pause between sentences
P3: pause within a sentence

© Hiroya Fujisaki
Application of the Model to Various Other Languages
## Languages Analyzed by the Current Author and His Colleagues

<table>
<thead>
<tr>
<th>Language</th>
<th>Authors and Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>Fujisaki et al. (1969 ~ )</td>
</tr>
<tr>
<td>Chinese (Mandarin)</td>
<td>Fujisaki, Hallé, Lei (1987 ~ )</td>
</tr>
<tr>
<td>English</td>
<td>Fujisaki et al. (1978 ~ )</td>
</tr>
<tr>
<td>German</td>
<td>Fujisaki and Mixdorff (1994 ~ )</td>
</tr>
<tr>
<td>Greek</td>
<td>Fujisaki et al. (1997 ~ )</td>
</tr>
<tr>
<td>Hindi</td>
<td>Fujisaki et al. (2003 ~ )</td>
</tr>
<tr>
<td>Korean</td>
<td>Fujisaki et al. (1996 )</td>
</tr>
<tr>
<td>Polish</td>
<td>Fujisaki et al. (2000 )</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Fujisaki et al. (2003 ~ )</td>
</tr>
<tr>
<td>Russian</td>
<td>Fujisaki et al. (2005 ~ )</td>
</tr>
<tr>
<td>Spanish</td>
<td>Fujisaki et al. (1994 )</td>
</tr>
<tr>
<td>Swedish</td>
<td>Fujisaki &amp; Ljungqvist (1993)</td>
</tr>
<tr>
<td>Thai</td>
<td>Fujisaki, Ohno &amp; Luksaneeyanawin (2002 ~ )</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Mixdorff, Fujisaki et al. (2003 ~ )</td>
</tr>
</tbody>
</table>

---

**Lanxi**  
Rickard (2005)
Application of the Model 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, Shanghainese, Suzhou, Wujiang), 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
- Korean
- German
- Polish
- Greek
- Spanish
Application of the Model to Languages other than Japanese

**Swedish**

Av invånarna i Sverige beräknas för närvarande cirka tio procent vara av utländsk härkomst.

**Hindi**

मैसी जी को वमाधि पर फूल चढ़ाने के लिए लोग रजिपाट जाते हैं।

**Portuguese**

Deseja escolher algum produto desta lista?
Application of the Model to Languages other than Japanese

Mandarin

Thai

British English

Where are you going?

Have you made up your mind?

You always try to do everything at the last minute.
Demonstration of $F_0$ Contour Synthesis (1)

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 (Mandarin), 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. Mandarin: 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.*
**Typological Grouping of Languages Based on Tone/Accent Command Polarity**

<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**, Bangla, 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 Novel Representation of the Tone Systems of Some Tone Languages
Background

• Traditional or conventional methods for describing tones do not have the accuracy and generality that are attained by the methods for describing segmental features.

• Articulatory processes for producing segments are observable, at least partially, and can be objectively described.

• Phonatory process for producing tones are much more difficult to observe directly, so that scholars in the earlier days had to rely almost exclusively on auditory perception in describing tones and their systems.

• We need a more reliable, objective method for tone description.
In the following section, we first review three major approaches and discuss their shortcomings:

1. The approach adopted by ancient Chinese linguists (since more than 1400 years ago)
2. The approach adopted by Western phonologists since last century
3. The approach started by Y. R. Chao
(1) Traditional Chinese Terminology

It is well-known that various dialects of Chinese have had tones at least since the Southern and Northern Dynasty (5th ~ 6th Century A.D.).

According to the traditional Chinese terminology, Middle Chinese (7th ~ 10th century A. D.) had four categories of tones:

平 (level), 上 (elevating), 去 (departing), and 入 (entering), each of which was divided into two subcategories (with two registers) according to the voicing property of the initial:

陰 (upper) and 阳 (lower) for voiceless and voiced initials, respectively,

and hence eight tones altogether.
These traditional tone names are still maintained in Modern Chinese (since 1920s), but in most of the dialects they do not reflect actual tonal features.

For example, the distinction between 阴 and 阳 is preserved in many southern dialects but not in Mandarin. Certain tones in Middle Chinese have been merged in various ways, while a few tones have been split to produce more tone categories.

Due to its historical origin, Vietnamese tone system preserves the tonal categories of Middle Chinese, though it groups 去 and 入 into one category. Hence, both Vietnamese and Chinese names of the tones are listed in the following Table A. Vietnamese tones are characterized not only by the pitch pattern but also by other features of voice production.
Table A. Descriptions of Tones of Mandarin, Shanghainese, Thai, Vietnamese and Lanxi

<table>
<thead>
<tr>
<th></th>
<th>Mandarin</th>
<th>Shanghai</th>
<th>Thai</th>
<th>Vietnamese</th>
<th>Lanxi</th>
</tr>
</thead>
<tbody>
<tr>
<td>平</td>
<td>陰平</td>
<td>H 55</td>
<td>H 45</td>
<td>H 44</td>
<td>M 33</td>
</tr>
<tr>
<td></td>
<td>阳平</td>
<td>R 35</td>
<td>R 14</td>
<td>LF 31</td>
<td>LF 21</td>
</tr>
<tr>
<td>上</td>
<td>陰上</td>
<td>L 21(4)</td>
<td>L 21</td>
<td>D 21(4)</td>
<td>HF 534</td>
</tr>
<tr>
<td></td>
<td>阳上</td>
<td>L 21(4)</td>
<td>L 21</td>
<td>HRG 32 4</td>
<td></td>
</tr>
<tr>
<td>去</td>
<td>陰去</td>
<td>F 51</td>
<td>F 41</td>
<td>HR 35</td>
<td>HR 45</td>
</tr>
<tr>
<td></td>
<td>阳去</td>
<td>F 41</td>
<td>F 41</td>
<td>LG 21</td>
<td>LG 21</td>
</tr>
<tr>
<td>入</td>
<td>陰入</td>
<td>H’ 5</td>
<td></td>
<td>HR’ 34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>阳入</td>
<td>L’ 2</td>
<td></td>
<td>LR’ 12</td>
<td></td>
</tr>
</tbody>
</table>

© Hiroya Fujisaki
(2) Notations of Tone in Modern Phonology

In modern phonology, a set of descriptive features is used to characterize both the height and the shape of pitch patterns of the tones, as shown in the left column for each language/dialect in Table A, where ‘H’, ‘M’ and ‘L’ indicate high, mid and low levels while ‘R’ and ‘F’ indicate rising and falling contours respectively. By combining these two sets of features, more complex tones can also be described.

In Table A, a single quotation mark is used indicate an entering tone to distinguish it from its non-entering counterpart. For Vietnamese, ‘D’ is used to indicate dipping (falling with an optional rising) tones, and ‘G’ to idicate glottalized tones.

However, by the addition of these somewhat ad hoc features, the consistency and parsimony of the whole system is no more guaranteed as a system of phonemic features.
(3) Y. R. Chao’s Tone Code System

Chao proposed a phonetically-oriented approach to represent the pitch patterns of lexical tones by a 5-level tone letter or a tone code system. It was later revised and was adopted by IPA.

The five levels (from 1 to 5), roughly correspond to a relative musical scale, and are used to indicate the relative pitches of low, mid-low, mid, mid-high and high, respectively. This system is used to describe the pitch pattern, not only of a level tone but also of a contour tone, by a string of numbers (at most four but usually not more than three numbers).

In spite of its greater flexibility for describing various tone features, this system also has limitations and problems.
Problems with Y. R. Chao’s Tone Code System

1. The five levels are originally defined perceptually and are subjective and relative. Although contemporary researchers try to apply these levels by visual inspection of measured $F_0$ contours, there is no well-defined convention on the procedure. Thus it is also subjective.

2. As a result, notations vary both among speakers of the same dialect and among researchers who identify and transcribe the tones, as exemplified by the next table by Yuk-Man Cheung of CityUHK.

3. This notation system is valid only for representing tones in isolated syllables.

4. It cannot characterize the continuous and dynamic nature of $F_0$ contours.
Comparison of tone values of Hakka Chinese reported by different researchers (Courtesy of Yuk-Man Cheung, CityUHK)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yin-Ping</td>
<td>[44]</td>
<td>[44]</td>
<td>[44]</td>
<td>[33]</td>
<td>[33]</td>
</tr>
<tr>
<td>Shang-Sheng</td>
<td>[31]</td>
<td>[31]</td>
<td>[31]</td>
<td>[31]</td>
<td>[31]</td>
</tr>
<tr>
<td>Qu-Sheng</td>
<td>[53]</td>
<td>[52]</td>
<td>[52] or [42]</td>
<td>[55]</td>
<td>[51]</td>
</tr>
<tr>
<td>Yin-Ru</td>
<td>[1]</td>
<td>[1]</td>
<td>[1] or [21]</td>
<td>[31]</td>
<td>[31]</td>
</tr>
</tbody>
</table>

Comparison between the narrow transcription for the six citation tones in this study* and the impressionistic descriptions in the past studies.

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>↓</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>阳平声(第二声)</td>
<td>高降</td>
<td>↑</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>全降</td>
<td>＼</td>
<td>51</td>
</tr>
</tbody>
</table>

The syllable “yi” uttered with four different $F_0$ contours, and its four different meanings.

© Hiroya Fujisaki
The polarity of tone commands for the four tones are: positive for T1, initially negative but switched to positive for T2, negative for T3, and Initially positive but switched to negative for T4. In this work, however, we assume that each tone has always two tone commands, one occurring early in the syllable, and another occurring late in the syllable. Thus T1 is regarded to possess two positive commands of an equal amplitude, while T3 is regarded to possess two negative commands of an equal amplitude.
Phonological Contrasts of Four Tones and the Light Tone of Standard Chinese in Terms of Tone Command Polarity

Command 2 (late in the syllable) +

Tone 2(R) (Light Tone) Tone 1(H)

Command 1 (early in the syllable) +

Tone 3(L) Tone 4 (F)
Analysis-by-Synthesis of Five Citation Tones of Shanghainese

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 80
Phonological Contrasts of Five Citation Tones of Shanghainese in Terms of Tone Command Polarity

Command 2
(late in the syllable)

Command 1
(early in the syllable)

T1*

T1

T1* is a variation of T1 in isolated syllables or at phrase end

(..) indicates entering tones

T3 (T5) (T2) T4
Tone Contours and the Underlying Commands of Thai

Waveform

\[ F_0 \text{ Contour} \]

Tone Command

Polarity of Commands

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 82
Phonological Contrasts of Five Tones of Thai in Terms of Tone Command Polarity

Command 2 (late in the syllable)

Command 1 (early in the syllable)

T2(R)  T1(H)  T0(M)  T3(L)  T4(F)
Raw mean $F_0$ plotted as a function of mean absolute duration for a male speaker’s seven citations tones of Lanqi

From:
K. Rickard,
*An acoustic-phonetic descriptive analysis of Lanqi citation tones.*
Phonological Contrasts of Seven Tones of Lanqi (蘭溪) in Terms of Tone Command Polarity

Command 2
(late in the syllable)

\[ + \]
\[ \uparrow \]

HR/HR’

- \[ LR/LR’ \] M HF +

Command 1
(early in the syllable)

LF

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27)
Phonological Contrasts of Six Tones of Vietnamese in Terms of Tone Command Polarity

Command 2
(late in the syllable)

Command 1
(early in the syllable)

ngã/sắc

ngang

hỏi/năng

huyễn

© Hiroya Fujisaki
Phonological Structures of Tone Systems of Five Tone Languages

(a) Mandarin

(b) Shanghainese

(c) Thai

(d) Vietnamese

(e) Lanqi

__: indicates the overlapping of two tones types, e.g., entering and non-entering tones.

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 87
Tone values of the six citation tones of Hakka*  
(Courtesy of Yuk-Man Cheung, CityUHK)

[u] for Female Speaker 2

Phonological structures of six tone systems of including Meixian Hakka

(a) Mandarin
(b) Shanghainese
(c) Thai
(d) Vietnamese
(e) Lanqi
(f) Meixian Hakka

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 89
Analysis-by-Synthesis of Nine Tones of Cantonese

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27)
Phonological Contrasts of Nine Tones of Cantonese in Terms of Tone Command Polarity

Command 1
(early in the syllable)

Command 2
(late in the syllable)

T1 (T7) +

T2

T3 (T8) -

T5 -

T6 (T9)

T4

(../) indicate entering tones

© Hiroya Fujisaki

TAL2012, Nanjing (2012-05-27) 91
Phonological structures of tone systems of six tone languages

(a) Mandarin
(b) Shanghainese
(c) Thai
(d) Vietnamese
(e) Lanqi
(g) Cantonese

© Hiroya Fujisaki
Tone Space with Three Tone Commands
Phonological Contrasts of Some Tones of Lhasa in Terms of Tone Command Polarity (1)
Phonological Contrasts of Eight Tones of Lhasa in Terms of Tone Command Polarity (2)
Discussion

Although an approximate correspondence can be observed between the proposed way of tone type representation and the traditional 5-level tone code system if we take the mid level 3 as a reference and map the higher and lower levels to positive and negative commands respectively, the essential difference lies in that a command in our model produces a dynamic $F_0$ contour instead of a static $F_0$ value through the tone control mechanism.

It is also to be noted that the actual amplitude of the tone commands are not constrained to discrete levels but can assume values on a continuum, and can vary depending on such para- and non-linguistic factors as emphasis and emotion, so that the use of five discrete levels for continuous $F_0$ contours is misleading (over-differentiation in phonology yet over-simplification in phonetics). We consider that the proposed ternary representation for the tone commands is sufficient for describing the phonological structure as far as these five tone languages (excluding Cantonese) are concerned.
Conclusions

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 – continued

4. 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 accuracy.

5. Differences in $F_0$ contours of languages originate mainly from differences in the patterns of commands for the local components.

6. Apart from its usefulness in phonetic studies and technological applications (e.g., high-quality speech synthesis), the model leads to a novel way of representing the phonological structure of a tone language in terms of timing and polarity of tone commands.
Dedication and Acknowledgment

- I wish to dedicate this small talk to my dear and respected friend, Prof. William S.-Y. Wang, whom I met for the first time in 1959 at MIT, especially because we share at least a part of our academic training and career, viz. as Professor of Electrical Engineering.

- We would like to thank Dr. Phil Rose of Australian National University for helpful discussions and for providing the $F_0$ data on Lanqi tones.

- We would also like to express our special thanks to Ms. Yuk-Man Cheung of City University of Hong Kong for her generosity in allowing us to use her data and results presented at TAL 2006 held in La Rochelle, France.
Thank you for your attention!

多謝聆聽！
Selected References (1)


© Hiroya Fujisaki
Selected References (2)


© Hiroya Fujisaki
Selected References (3)


Selected References (4)


Note: Some of our recent papers on Cantonese are not yet included in the above list.