Comparison of Declarative and Interrogative Intonation in Chinese

Jiahong Yuan, Chilin Shih and Greg Kochanski

(Cornell University) (Bell Laboratories – Lucent Technologies)

Lucent Technologies
Bell Labs Innovations

Abstract

In most of the world's languages, one can transform a statement into a question by raising the pitch at the end. This transformation is dubious in tone languages, as it could transform one lexical item into another.

- How does one ask question in a tone language?
- Is there is a question phrase curve?
- Is there a question boundary tone?

We build and train models of Mandarin Chinese intonation to answer these question. The resulting models have RMS errors of 10 Hz, or 1 semitone.

We find that questions are marked by:

- More careful intonation, and a greater range of f_0 at the end of the sentence.
- A slightly raised, but otherwise unremarkable, phrase curve.

What is Stem-ML?

Stem-ML combines several ideas:

- People plan their utterances several syllables in advance.
- People produce speech that is optimized to meet their needs:
  - Speech is a balance between accurate communication and ease of production.
  - People can practice all tonal combinations.
- A simple, but reasonable model for the dynamics of the muscles that control pitch.
- The concept of a strength associated with each syllable:
  - High strength → nearly ideal shape.
  - High strength ∼ careful articulation.
  - High strength ∼ expanded pitch range

How does Stem-ML work?

Stem-ML calculates a pitch curve by finding the curve that minimizes *effort+error*, where *effort* approximates the physiological effort: it is zero if muscles are stationary, and increases as motions become faster and stronger. The *error* term measures how far the pitch curve deviates from the ideal template.

Equations have been simplified for presentation.

$$effort = \sum_t \dot{p}_t^2 + \text{smooth}^2 \ddot{p}_t^2 \qquad (1)$$

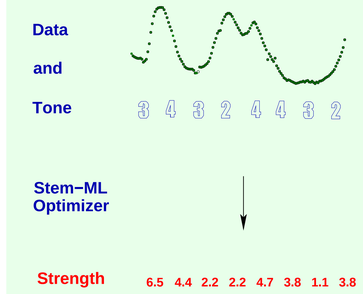
(p_t is the pitch, $\dot{p}$ is the rate of change, and "smooth" is a speaker-dependent constant).

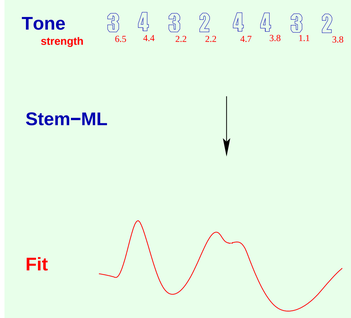
$$error = \sum_{k \in \text{tags}} s_k^2 r_k \qquad (2)$$

where (y is the template, and p is the pitch, and s_k is the strength of syllable k).

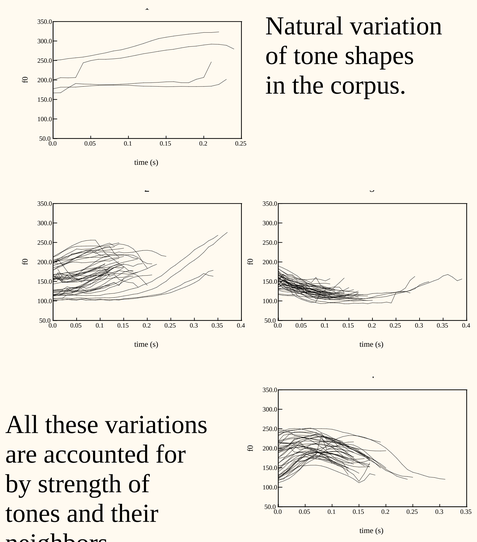
$$r_k = \sum_{t \in \text{tag } k} (p_t - y_t)^2 \qquad (3)$$

(This is the error summed over a syllable).



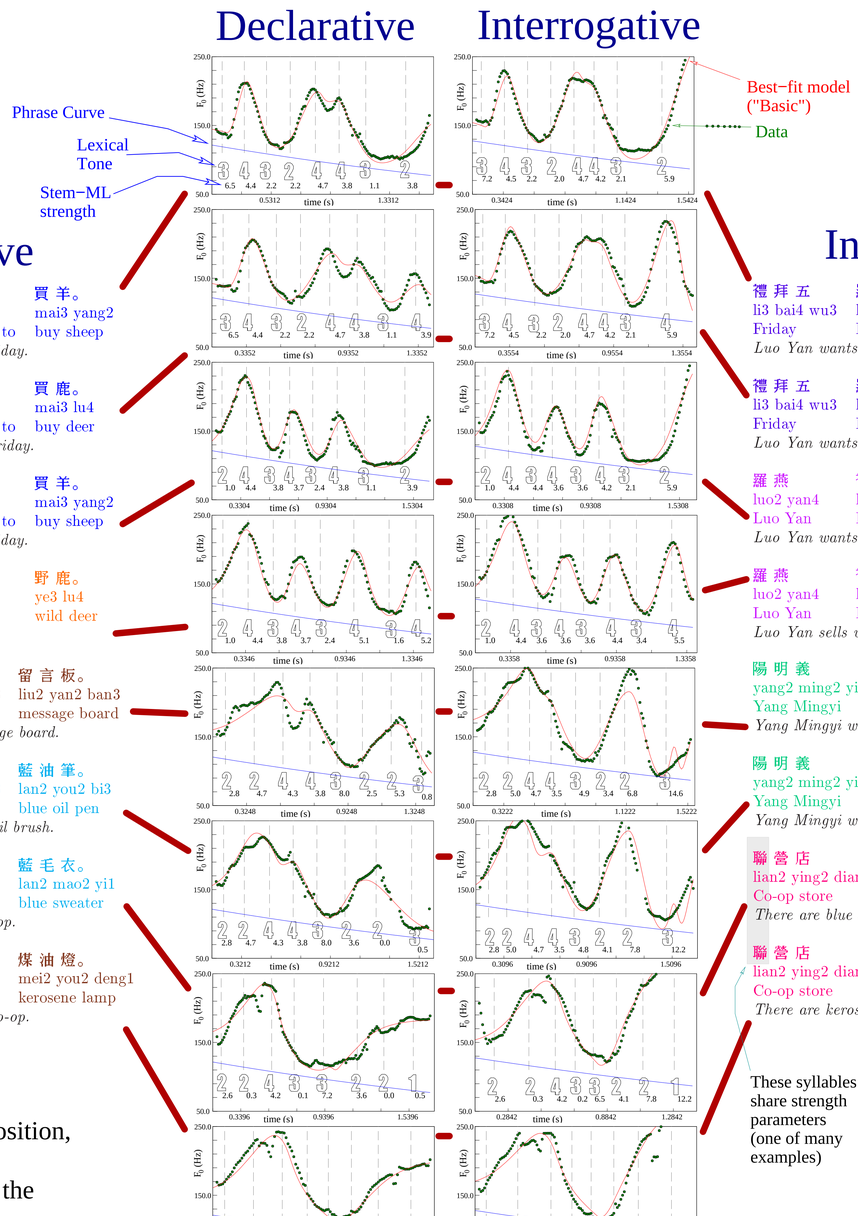


- Corpus
  - 16 sentences: paired declaratives and questions
    (The questions have a faintly incredulous flavor, in Mandarin as well as English.)
- Design:
  - Avoid question words to get the same tonal and segmental sequences in sentence pairs
  - Natural, readable sentences
  - Mostly sonorant sounds
  - Various tonal combinations
  - Varying word boundary locations
- Data preparation:
  - F_0 extraction with ESPS/Waves
  - Manual error correction
  - Manual labels of syllable boundaries
  - Manual labels of lexical tones



Natural variation of tone shapes in the corpus.

All these variations are accounted for by strength of tones and their neighbors.



Declarative

禮 拜 五 li3 bai4 wu3 Friday
羅 燕 luo2 yan4 Luo Yan
要 yao4 want to
買 羊。 mai3 yang2 buy sheep
Luo Yan wants to buy sheep on Friday.

禮 拜 五 li3 bai4 wu3 Friday
羅 燕 luo2 yan4 Luo Yan
要 yao4 want to
買 鹿。 mai3 lu4 buy deer
Luo Yan wants to buy a deer on Friday.

羅 燕 luo2 yan4 Luo Yan
禮 拜 五 li3 bai4 wu3 Friday
要 yao4 want to
買 羊。 mai3 yang2 buy sheep
Luo Yan wants to buy sheep on Friday.

羅 燕 luo2 yan4 Luo Yan
禮 拜 五 li3 bai4 wu3 Friday
賣 mai4 sell
野 鹿。 ye3 lu4 wild deer
Luo Yan sells wild deer on Friday.

陽 明 義 yang2 ming2 yi4 Yang Mingyi
要 yao4 want to
買 mai3 buy
留 言 板。 liu2 yan2 ban3 message board
Yang Mingyi wants to buy a message board.

陽 明 義 yang2 ming2 yi4 Yang Mingyi
要 yao4 want to
買 mai3 buy
藍 油 筆。 lan2 you2 bi3 blue oil pen
Yang Mingyi wants to buy a blue oil brush.

聯 營 店 lian2 ying2 dian4 Co-op store
裡 li3 inside
有 you3 has
藍 毛 衣。 lan2 mao2 yi1 blue sweater
There are blue sweaters in the Co-op.

聯 營 店 lian2 ying2 dian4 Co-op store
裡 li3 inside
有 you3 has
煤 油 燈。 mei2 you2 deng1 kerosene lamp
There are kerosene lamps in the Co-op.

Parameter sharing: words in the same position, with the same color are assumed to have the same strength.

Interrogative

禮 拜 五 li3 bai4 wu3 Friday
羅 燕 luo2 yan4 Luo Yan
要 yao4 want to
買 羊？ mai3 yang2 buy sheep
Luo Yan wants to buy sheep on Friday?

禮 拜 五 li3 bai4 wu3 Friday
羅 燕 luo2 yan4 Luo Yan
要 yao4 want to
買 鹿？ mai3 lu4 buy deer
Luo Yan wants to buy a deer on Friday?

羅 燕 luo2 yan4 Luo Yan
禮 拜 五 li3 bai4 wu3 Friday
要 yao4 want to
買 羊？ mai3 yang2 buy sheep
Luo Yan wants to buy sheep on Friday?

羅 燕 luo2 yan4 Luo Yan
禮 拜 五 li3 bai4 wu3 Friday
賣 mai4 sell
野 鹿？ ye3 lu4 wild deer
Luo Yan sells wild deer on Friday?

陽 明 義 yang2 ming2 yi4 Yang Mingyi
要 yao4 want to
買 mai3 buy
留 言 板？ liu2 yan2 ban3 message board
Yang Mingyi wants to buy a message board?

陽 明 義 yang2 ming2 yi4 Yang Mingyi
要 yao4 want to
買 mai3 buy
藍 油 筆？ lan2 you2 bi3 blue oil pen
Yang Mingyi wants to buy a blue oil brush?

聯 營 店 lian2 ying2 dian4 Co-op store
裡 li3 inside
有 you3 has
藍 毛 衣？ lan2 mao2 yi1 blue sweater
There are blue sweaters in the Co-op?

聯 營 店 lian2 ying2 dian4 Co-op store
裡 li3 inside
有 you3 has
煤 油 燈？ mei2 you2 deng1 kerosene lamp
There are kerosene lamps in the Co-op?

These syllables share strength parameters (one of many examples)

These syllables share strength parameters (one of many examples)

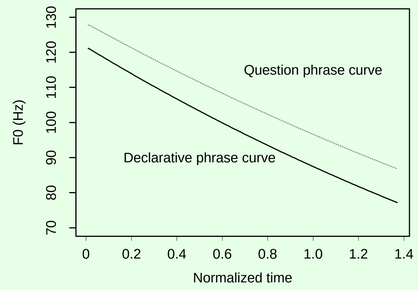
Basic Model

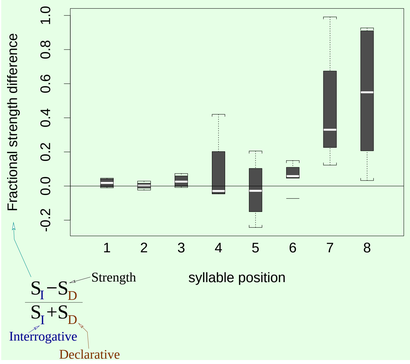
- Intonation is generated by just 5 templates (one for all tone 1 syllables ... one for tone 4, neutral tone). The templates are stretched ($\longleftarrow$ *time* $\longrightarrow$) and scaled ($\Uparrow$ *pitch* $\Downarrow$) for each syllable.
- Pitch scaling and the Stem-ML ***strength*** are tied together.
- Among declarative sentences, all equivalent words (see color code) share the same strength. Likewise for questions.
- All declarative sentences share one straight-line phrase curve, questions share another.
- All utterances share five speaker-dependent parameters.

This yields a model with 93 free parameters, or 0.7 per syllable, or 1 per 0.25 seconds.

Phrase curves

The best-fit shows that questions have somewhat raised pitch, and a similar declination rate to declarative sentences.





A Rising tail?

The "Basic" model fits well, but it has strong assumptions. Let us relax the assumption that a phrase curve is a straight line, because many people would expect the phrase curve to

- rise at the end for questions, and
- fall at the end for statements.

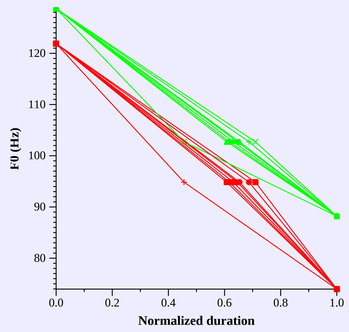
We test this by adding a free parameter to each sentence's phrase curve so it consists of two linear segments. Each sentence has its own knot position, and the interrogative phrase curve is completely unconnected to the declarative phrase curve.

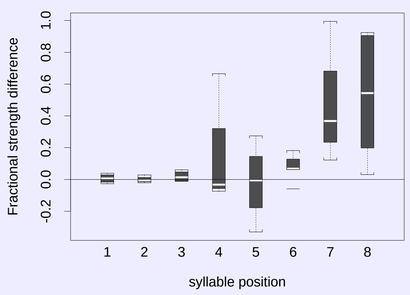
This yields a model with 103 free parameters, or 0.8 per syllable, or 1 per 0.23 seconds.

Phrase curve.

The best-fit phrase curve shows no evidence of a consistently different tail.

It shows a consistent shift, much as is seen in the "basic" model.





A Boundary Tone?

Let us extend the basic model to include a final boundary tone: boundary tones might be expected, because they are useful in describing languages like English.

We test this by adding an additional Stem-ML stress tag near the end of the sentence. We used a tag that has the ability to fix the pitch to a particular value relative to the phrase curve.

The boundary tones for all declarative sentences share one strength, and the interrogative sentences shared another.

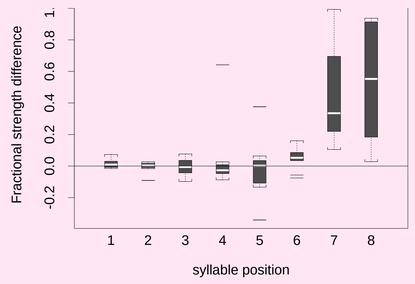
We did two tests, one with the extra tag at the very end, and one with it near the center of the last syllable.

This yields a model with 95 free parameters, or 0.7 per syllable, or 1 per 0.25 seconds.

Boundary Tone Strength Values

The strengths of the best-fit boundary tone in these two models are just 3% of the average strength of the lexical tones. This corresponds to a change of about 10 Hz, out of the typical 100 Hz pitch swings.

The boundary tones, as implemented here, are not an important part of the intonation.



All four models show a consistent pattern of prosodic strength: questions are stronger near the end. In the Stem−ML models we used here, the increase in strength causes the tones to match their templates more closely. The stronger templates have an expanded pitch range.

Conclusions

- How does one ask a question in a tone language?
  Increase the prosodic strength of the tones near the end of the sentence. This gives wider pitch swings and more careful intonation.
- Do questions have a different phrase curve?†
  The question phrase curve has a higher pitch than the declarative phrase curve.

  Contrary to previous descriptions, the two phrase curves are nearly parallel.

  The question phrase curve does not have a rising tail.
- Is there a question boundary tone?
  This model does not require one. Boundary tones add complexity without improving the fit.
- How does the phrase curve interact with lexical tones?†
  It raises the pitch of the sentence, but tone shapes are preserved.

† These phrase curves may just be a way to represent the average pitch and declination. We have no evidence for more complex phrase curves.

Q&A

Q: What does this model contribute to the understanding of prosody?

A:
- It puts the calculation of f_0 from tones on a firm footing, based on physiology and communication theory.
- It allows linguistic theories to be tested quantitatively.

Q: How can it work with so little data?

A: Stem-ML contains the basic rules for tonal coarticulation. Therefore, the system doesn't need to learn how each tone interacts with every other, just a few parameters that control the interactions.

Q: Where do the residual errors come from?

A:
- We assume that sets of several words share the same strength. This is just an approximation to the real complexity of language.
- Stem-ML ignores segmental effects,
- It uses a simplified model for muscle dynamics and laryngeal oscillations.

Q&A

Q: How does Stem-ML differ from most machine learning models of prosody?

A:
- It uses only 5 tone categories, lexically determined.
- The parameters can be understood and interpreted.
- The results can be portable to other problems.
- Models can be built with a very small data base.

Q: How does it differ from C-ToBI?

A:
- It uses lexical tone information to explain the f_0 curve.
- It can be automatically fit to the data.
- It can reconstruct the f_0 curve.
- It explains the differences between citation and actual tone shapes.

For more information, see
http://prosody.multimedia.bell-labs.com