Intonational tunes in imitative speech: Are phonological contrasts maintained?

Jeremy Steffman & Jennifer Cole
Northwestern University
MidPhon 2021 – September 18, 2021
jeremy.steffman@northwestern.edu
Introduction

• Prevalent theory of Intonational Phonology: intonational contours are phonologically specified as high (H) and low (L) tone targets

• Phonological inventories consist of tones which link to prominent positions and domain edges
American English: Nuclear Tunes

The final (nuclear) pitch accent + boundary tones = the nuclear tune, which may be composed of…

- 2 monotonal pitch accents: H*, L* (ignoring other pitch accents)
- 2 phrase accents: H-, L- (smaller phrase edge)
- 2 boundary tones: H%, L% (larger phrase edge)

8 (2 x 2 x 2) possible nuclear tunes composed of the above
The present study

Though 8 distinct tunes are predicted, we lack clear empirical evidence for an 8-way distinction

- Esp. from naïve participants - i.e. not trained intonational phonologists
- Some distinctions (e.g., rise vs. fall) are well studied - others not
- Some nuclear tunes are barley attested in labeled corpora (Dainora 2009)

Research question:
Do speakers evidence a robust 8-way distinction in nuclear tune shape, as predicted by the tonal inventory?
Do speakers vary in this regard?
Method: Imitative speech production paradigm

• Method from Chodroff & Cole (2019)
• Participants hear auditory models with resynthesized f0 which exemplify a tune

She quoted Helena

• They reproduce the tune on a new sentence and are instructed to do so in a way “that is familiar to you”
Method: Imitative speech production paradigm

Model sentences

She quoted Helena
Her name is Marilyn
He answered Jeremy

Target sentence
They honored Melanie
Method contd.

• 30 participants, 144 trials varying model sentence and speaker order

• **8 model nuclear tunes**
  • based on straight-line approximations in ToBI training materials
    (MIT OpenCourseWare: Veilleux et al., 2006 - based on Pierrehumbert, 1980)
  • 5 target heights spread across model speakers’ pitch range
  • “preamble” f0 declines until nuclear region (same for all tunes)
  • nuclear region: always a trisyllabic, initial-stress, name

Her name is Marilyn
Model tunes: nuclear region

Nuclear tune, with lines indicating e.g., 'Hel-e-na'
Measures

• Time-normalized f0
  • For the nuclear region only
  • Measured via STRAIGHT, using VoiceSauce (Kawahara et al. 2005; Shue 2010)
  • 30 samples per word
  • Converted to ERB, and centered on speaker mean values
Data: speaker means

speaker-centered f0 (ERB)

HHH | HHL | HLH | HLL

LHH | LHL | LLH | LLL

prop. word duration
Analysis

K-means clustering for longitudinal data (Genolini et al. 2015)

- Assigns trajectories to clusters, and iteratively optimizes cluster membership to minimize within-cluster variation
- We tested 2 through 10 clusters, and assessed what number of clusters best partitions the data (Calinski-Harabasz Criterion)
Results
Input: unlabeled speaker means
Output: 5 clusters
Output: 5 clusters
Output: 5 clusters
Output: 5 clusters

- Cluster A
- Cluster B
- Cluster C
- Cluster D
- Cluster E

Cluster distribution:
- LLL: 0.13 0.07 0
- LLH: 0.97 0 0
- LHL: 0.07 0 0
- LHH: 0.07 0 0
- HLL: 0 0.97 0
- HLH: 0.03 0.97 0
- HHL: 0 0 0.77
- HHH: 0 0 0.73
Output: 5 clusters
Output: 5 clusters
Output: 5 clusters
What about individual speakers?

- Are any speakers capturing all distinctions?
- Is there variation?

To address this: clustering for each speaker on individual trajectories (not speaker means: 144 trajectories per speaker)
  - What are the optimal number of clusters for each speaker?
Individual clustering: some speakers resemble the group data.
Individual clustering: many speakers do not resemble the group data.
Individual clustering: a prevalent 2 cluster “rising”/“non-rising” dichotomy
Individual clustering: distribution of optimal clustering solutions
Conclusions

Research question:
Do speakers evidence a robust 8-way distinction in nuclear tune shape, as predicted by the tonal inventory?
Do speakers vary in this regard?

• Group level data: evidence for 5 distinct tune shapes

• Individual level data
  • Lots of variation
  • Many speakers are best characterized as having a rising/non-rising distinction, with various partitions of the 8 tunes into these categories
Take home message

• Evidence for a **hierarchy of distinctions**
  • all speakers differentiate some tunes from one another, some make additional distinctions
  • not predicted by the current AM theory

• Some *types* of differences are clearly more easily accessible to speakers
  • (in this paradigm, in the absence of context)
Many thanks!

We are grateful to

• The NSF
• The Northwestern ProSD Lab
• Stefanie Shattuck-Hufnagel
• Lisa Cox
References


Appendix: The 6 and 7 cluster speakers
Appendix: Calinski Harabatz Criterion

- Calinski-Harabasz criterion: the ratio of between cluster dispersion to within cluster dispersion (higher values = better)

\[
\frac{\text{dispersion}_{\text{between}}}{\text{dispersion}_{\text{within}}} \times \frac{N - k}{k - 1} = CH
\]

- Where
  - \(N\) is the number of vectors (trajectories)
  - \(k\) is the number of clusters
Appendix: Calinski Harabasz Criterion

• dispersion is computed with the time series vectors via matrices:

\[ B = \sum_{m=1}^{k} n_m (\overline{y}_m - \overline{y})(\overline{y}_m - \overline{y})^T \]

\[ W = \sum_{m=1}^{k} \sum_{l=1}^{n_m} (y_{ml} - \overline{y}_m)(y_{ml} - \overline{y}_m)^T \]

• Where

- \( n_m \) the number of trajectories in cluster \( m \)
- \( \overline{y}_m \) is the mean of the trajectories in cluster \( m \)
- \( \overline{y} \) is the mean of all the trajectories
- \( v^T \) is the transposition of vector \( v \)
Appendix: Calinski Harabasz Criterion

• Finally, the trace of covariance matrix $B$ and $W$ is taken (summing the diagonal coefficients in each matrix)

\[
\frac{\text{trace}(B)}{\text{trace}(W)} \times \frac{N - k}{k - 1} = CH
\]