Temporal Location of Perceptual Cues for Cantonese Tone Identification

Zoe Wai-Man Lam, Kathleen Currie Hall and Douglas Pulleyblank

Department of Linguistics
University of British Columbia

Outline of the presentation

1. Background
2. Research question
3. Methodology
4. Results
5. Discussion
6. Conclusion

F0 variation in language

Fundamental frequency (F0): the frequency of vocal fold vibration (in Hz)

F0 variation takes place within a **temporal window**.

- Lexical tone: temporal window = duration of a syllable
- Intonation: temporal window = duration of an utterance

Acoustic vs. perceptual

- **F0** is an ***acoustic*** term: the physical properties of the signal
- **Pitch** is a ***perceptual*** term: the hearer's perception of the signal
- **Tone** is a ***linguistic*** term: phonological categories to distinguish words

(Yip 2002)

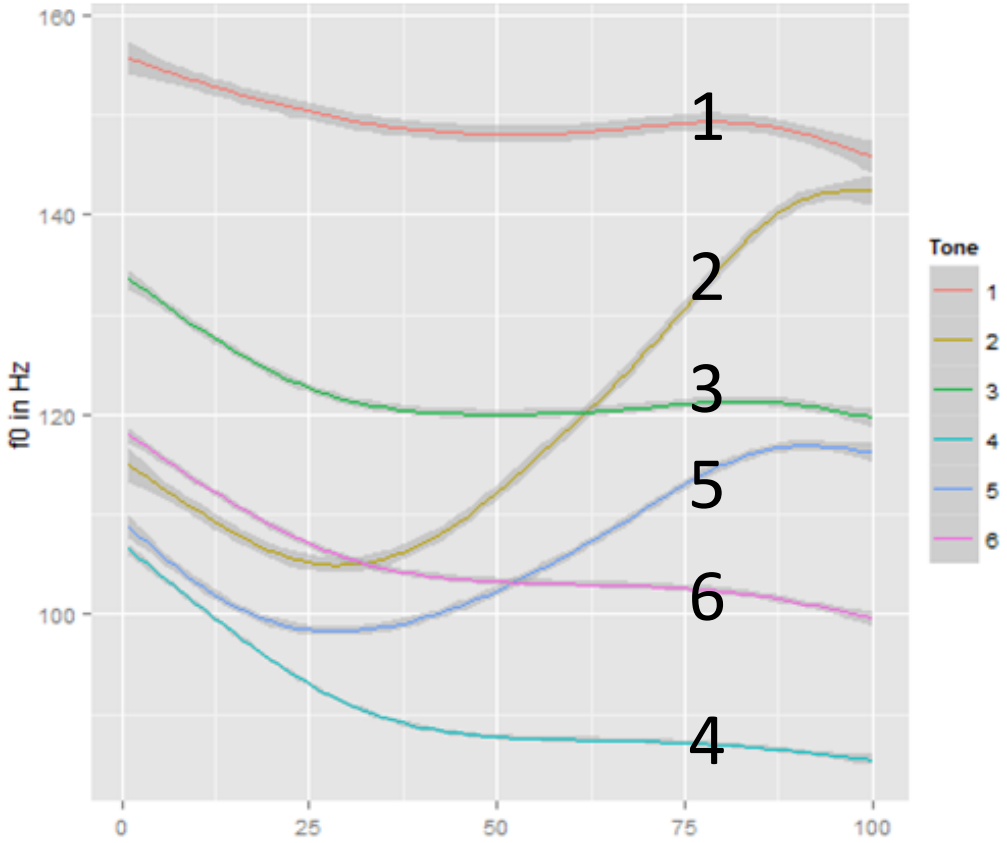
- Example: Yoruba has 3 level tones (H M L)
- In a disyllabic noun $\sigma_1\sigma_2$:

F0 of σ_2	**Perceived tone of σ_2**
Falling	Low tone
Flat	Mid tone

(Hombert 1976)

- Acoustic cues ≠ perceptual cues

Phonemic inventory of Cantonese tones



6 lexical tones

Tone 1	High	**level**
Tone 2	High	**rising**
Tone 3	Mid	**level**
Tone 4	Low	**falling***
Tone 5	Low	**rising**
Tone 6	Low	**level**

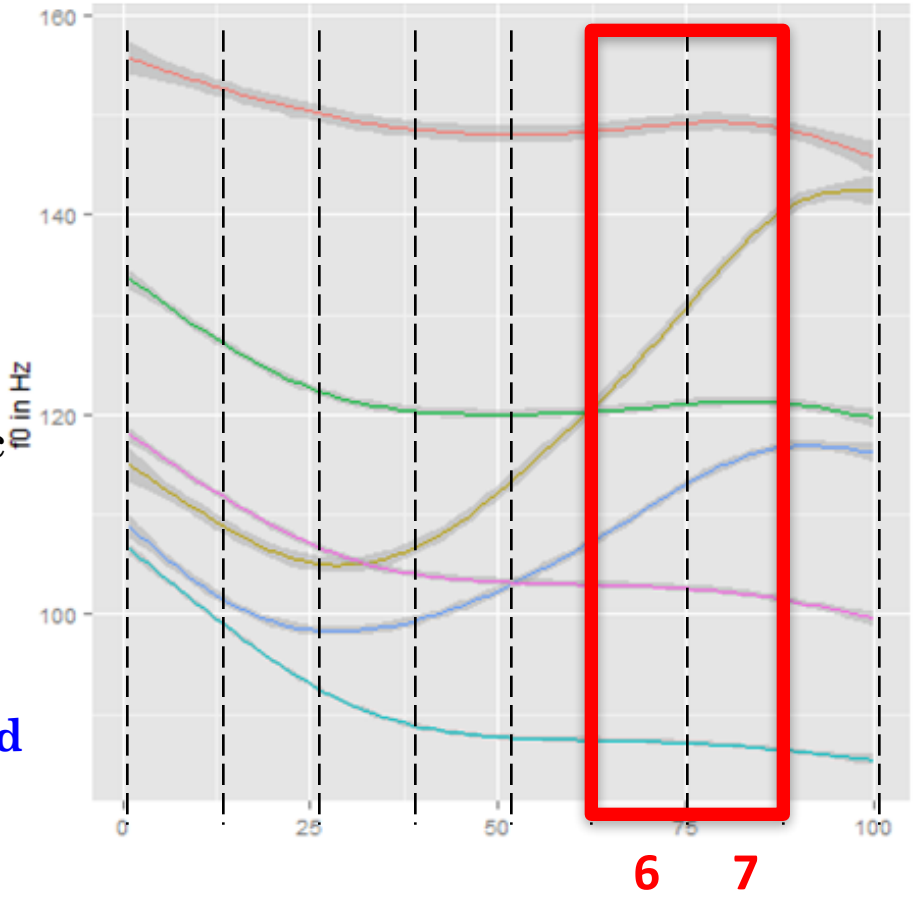
3 **level** tones (1, 3, 6)
3 **contour** tones (2, 4, 5)

*Slope of Tone 4 is similar to that of a level tone phonetically

Previous studies on tonal production

Khouw & Ciocca (2007)

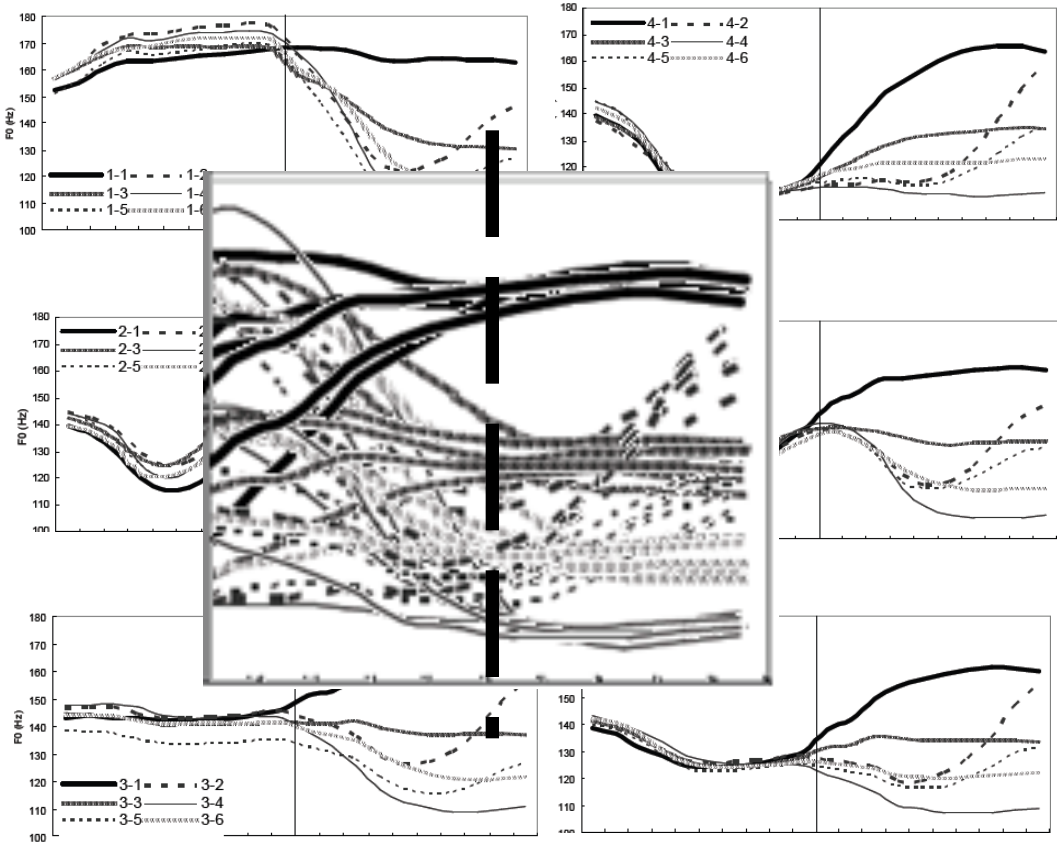
- Focus on syllables produced in isolation
- Goal: investigate which parts of the vocalic segment provide F0 information that is most strongly related to the identification of Cantonese tones
- F0 measurements are made for eight consecutive sections of the whole vocalic segment for each syllable.
- Discriminant analysis shows that the most important correlates of tone identity are in the later part of the vocalic segment, specifically the **6th and 7th sections**.



Previous studies on tonal production

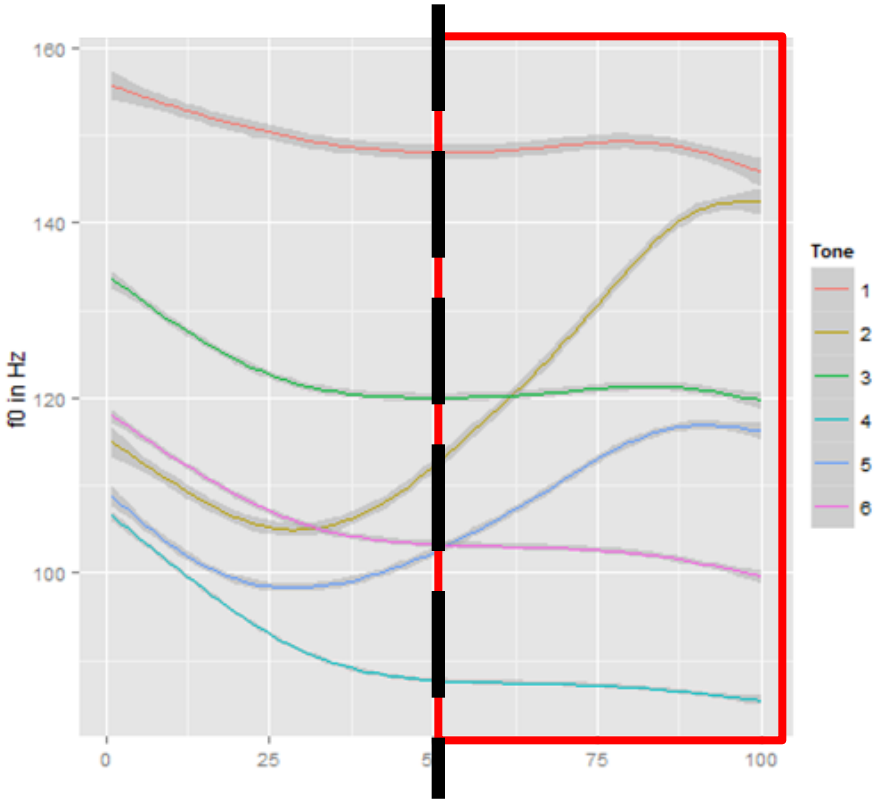
Wong (2007)

- Focuses on tonal coarticulatory effects in disyllabic sequences in a carrier sentence
- *lau lau* as target syllables
- Carryover effects are so strong that F0 transition from the tone of σ_1 to σ_2 takes up **50% of σ_2**, resulting in a great magnitude of F0 variation across different tonal contexts



Research question

Do native Cantonese speakers rely particularly on the perceptual cues from the latter portion of the tone for identification?



Our hypothesis

Stimulus preparation

- 4 naïve native speakers: 2 male, 2 female; age range 19-28
- Carrier sentence

  jau5 go3 giu3 $\underline{\sigma_1\sigma_2}$ ge3 je5
  haveCL call __ __ of thing
  'There is something called __ __.'
- $\sigma_1\sigma_2$ is any disyllabic combination of the following:

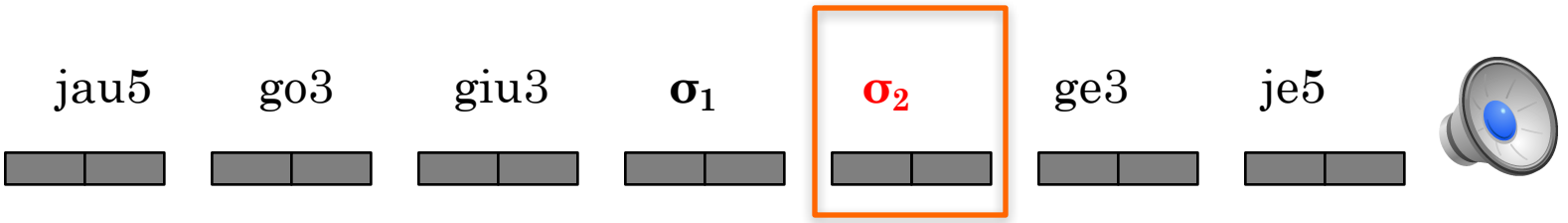
			si	se	fu
1	high level	55	詩 Poetry	些 Some	呼 To exhale
2	high rising	25	史 History	寫 To write	苦 Bitter
3	mid level	33	試 To try*	瀉 Diarrhea	富 Wealth
4	low falling	21	時 Time	蛇 Snake	符 To match
5	low rising	23	市 Market*	社 Society	婦 Woman/wife
6	low level	22	是 Right	射 To shoot	負 To load

- 1152 sentence tokens, divided into 8 groups of stimuli

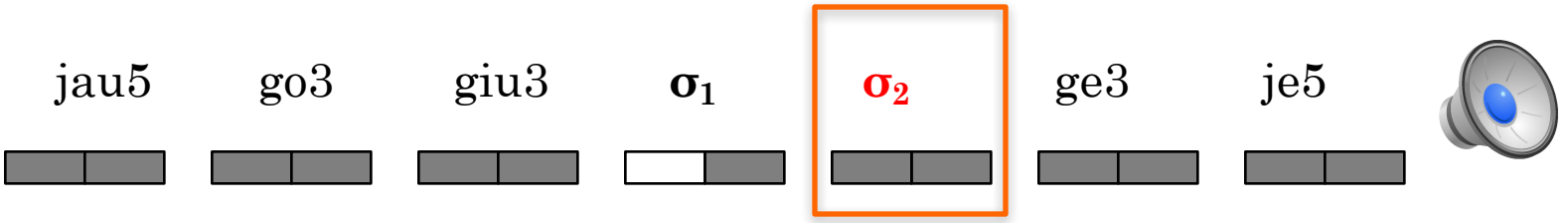
Four stimulus types

Control stimuli

Type A: Not manipulated



Type B: σ_2 not affected, but there is some disruption elsewhere



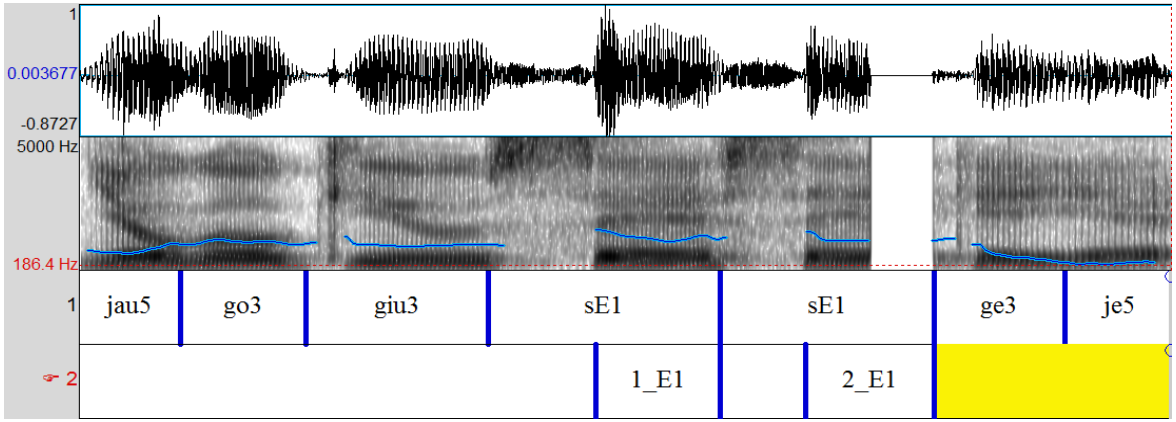
Four stimulus types

Type C: 1st half of the tone of σ_2 zeroed out

jau5 go3 giu3 σ_1 **σ_2** ge3 je5

Type D: 2nd half of the tone of σ_2 zeroed out

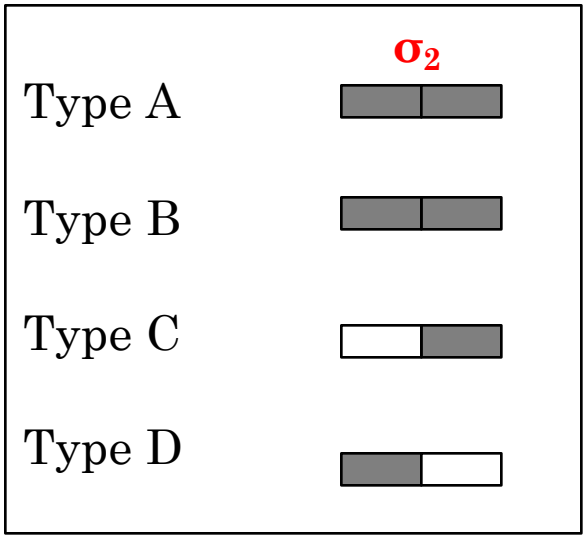
jau5 go3 giu3 σ_1 **σ_2** ge3 je5



Zeroing out done by a Praat script

Boundaries of the vocalic portion of the syllable were marked by the first author manually

Four predictions



Based on the hypothesis that the second half of the tone contains the crucial perceptual cues for tone identification, we predict that:

1. Type A = Type B
2. Type A = Type C
3. Type A > Type D
4. Type C > Type D

Perception study: Procedures

- Participants were presented with one sentence on the screen in each trial.
- They listened to an audio stimulus at the same time.
- **They were always asked what σ_2 was.**
- They were instructed to respond by pushing 1,2,3,4,5 or 6 on the keyboard.

有個叫優__嘅嘢

空格內的應是什麼字?
請按鍵盤上的 1,2,3,4,5或6。

1	2	3	4	5	6
優	柚	幼	油	友	又

Perception study: Procedures

Practice set

- 12 trials, using the syllable *jau,* produced by the first author
- All stimuli in the practice set were not manipulated (i.e. belong to Type A).
- Feedback was provided to the participants during the practice period.
- Data of participants who failed to get **at least 9 correct responses** were excluded from analysis.

Actual experiment

- 144 trials per participant, using *si*, *se* or *fu* syllables
- Randomized, not blocked according to stimulus type
- Participants were told that it's normal if some parts of the recording are not clear. They just needed to try their best to choose an answer.
- Feedback was NOT provided.

Perception study: The participants

- 63 participants from the UBC community participated in the study, who were either paid or given course credit.

- The data of 39 participants were excluded from the analysis.
  - Not a speaker of the Hong Kong variety of Cantonese (e.g. Malaysia)
  - Got fewer than 9 correct answers out of 12 Qs in the practice set

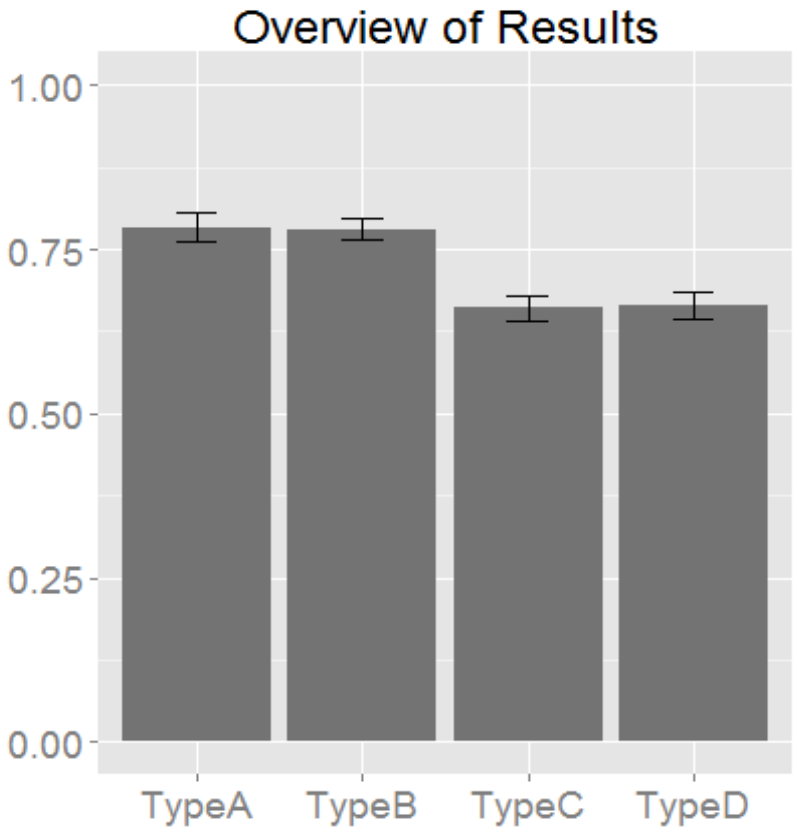
- The data of **24** participants were included in the analysis.
  - Got at least **9 correct answers** out of 12 Qs in the practice set
  - **Self-rated 6 or 7** out of 7 for their Cantonese proficiency
  - **Able to read** Traditional Chinese characters, **had spent at least 14 years in Hong Kong**
  - Subject 107 and 501 have never lived in Hong Kong, but their parents are from Hong Kong; results similar to other participants
  - Age: 18-36

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Results (Version 1)

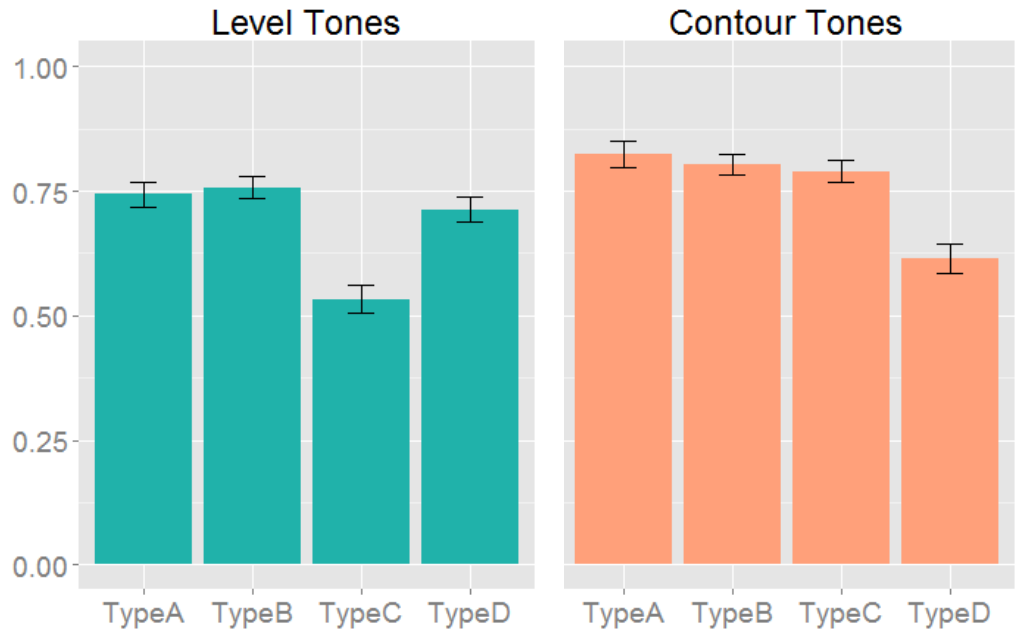
Are these predictions correct?

1. Type A = Type B Yes
2. Type A = Type C No
3. Type A > Type D Yes
4. Type C > Type D No



Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Results (Version 2)



Are these predictions correct?

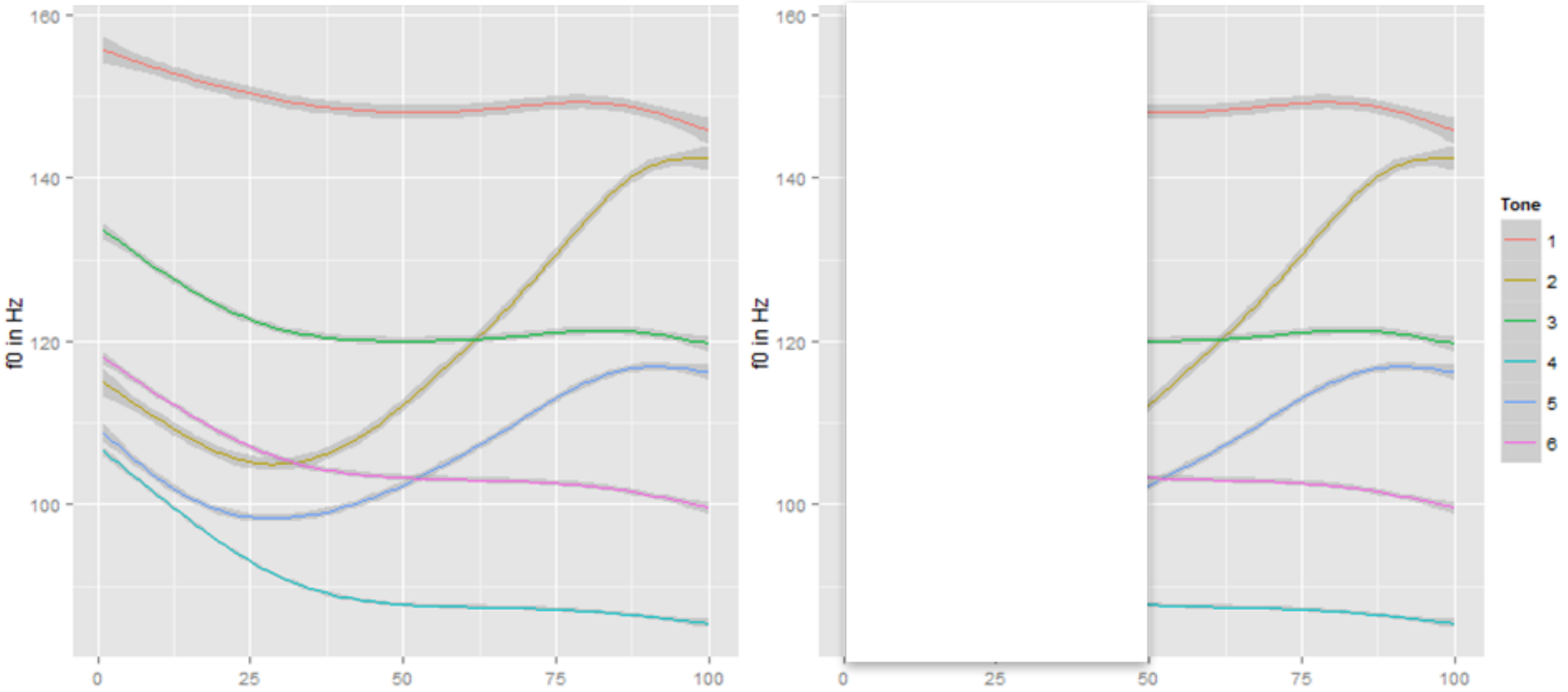
1. Type A = Type B
2. Type A = Type C
3. Type A > Type D
4. Type C > Type D

Yes only for contour tones but not level tones

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

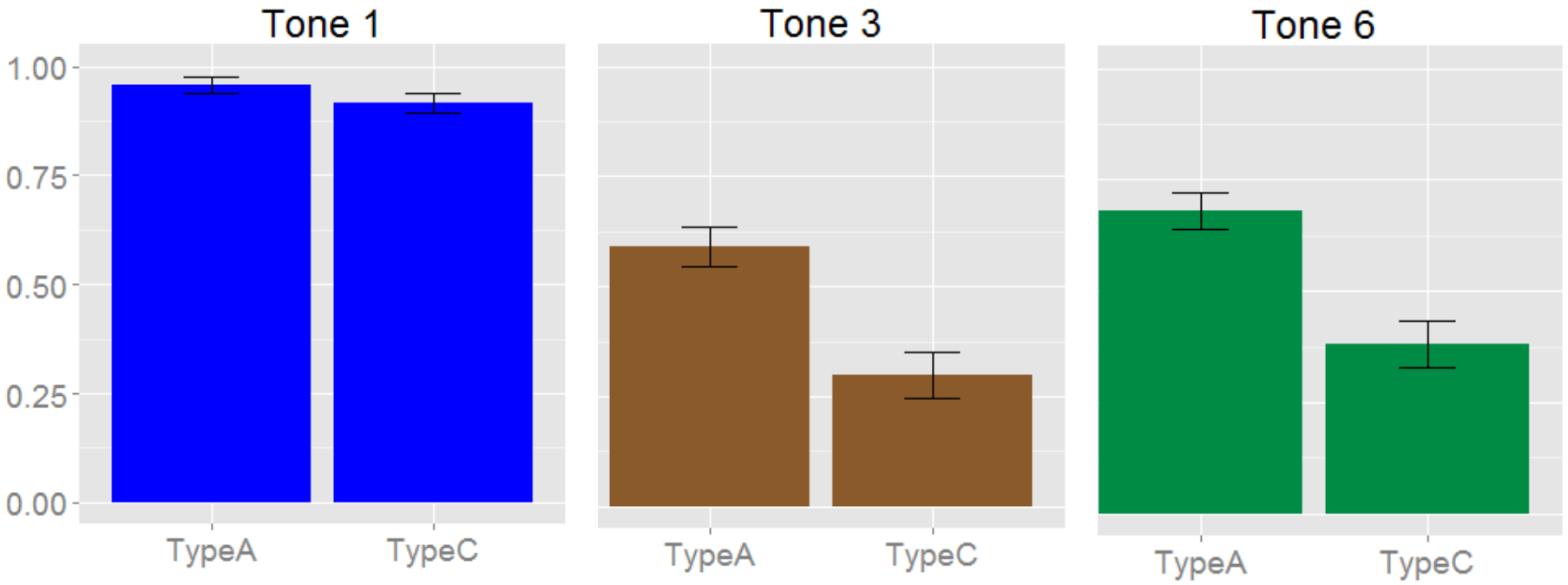
Prediction

Type A = Type C



Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

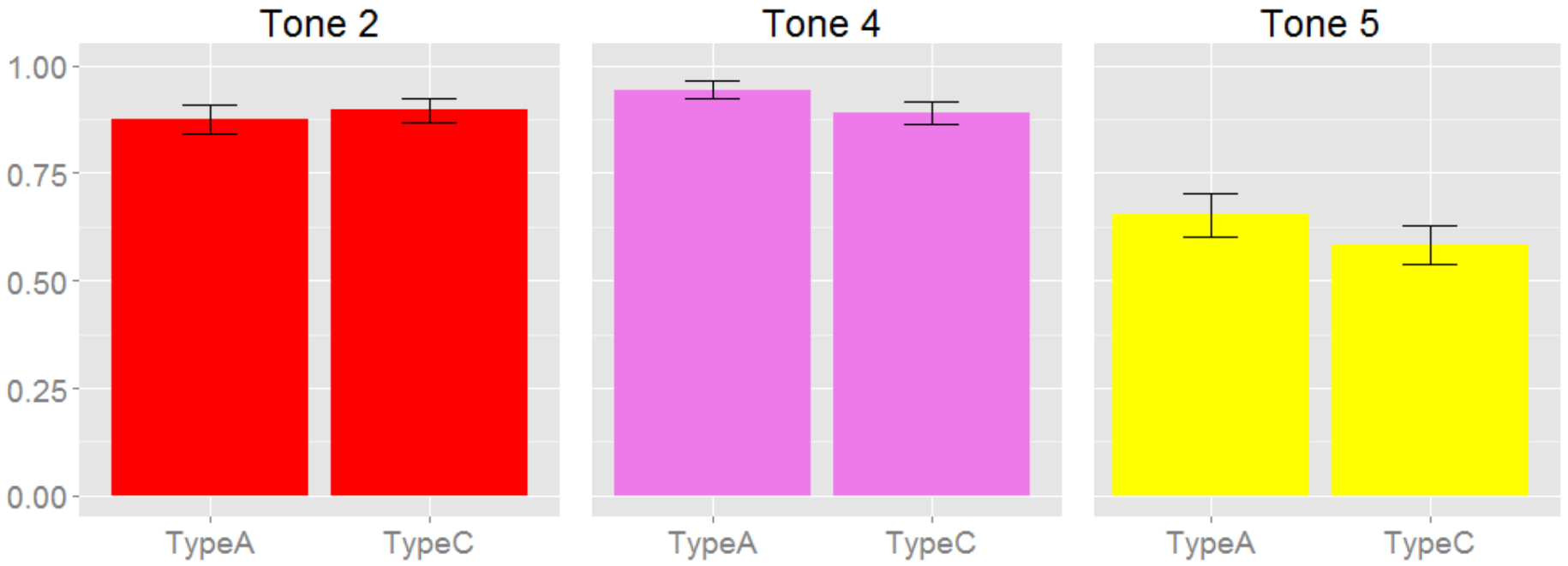
Level tones: Type A (control) vs Type C (1st half removed)



Prediction: Type A = Type C ✗

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Contour tones: Type A (control) vs Type C (1st half removed)

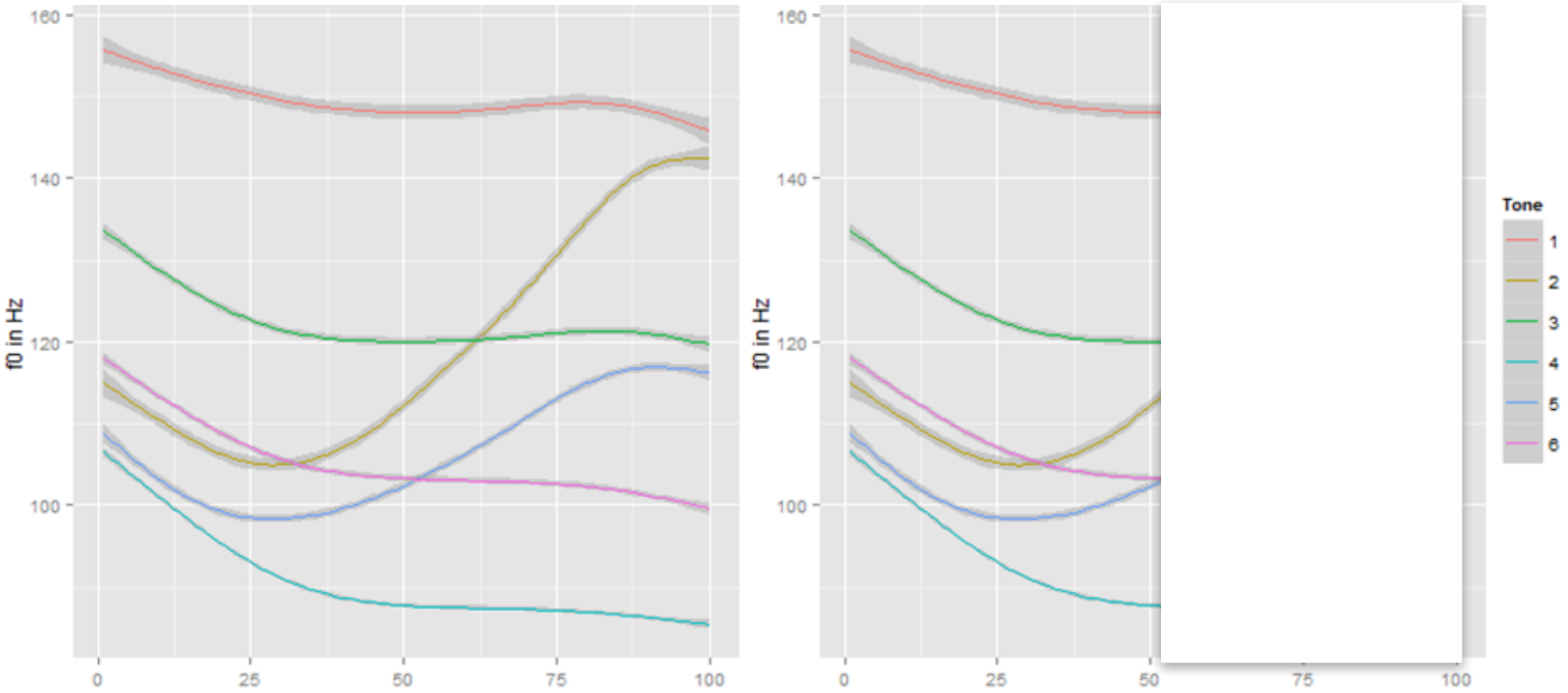


Prediction: Type A = Type C ✔

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

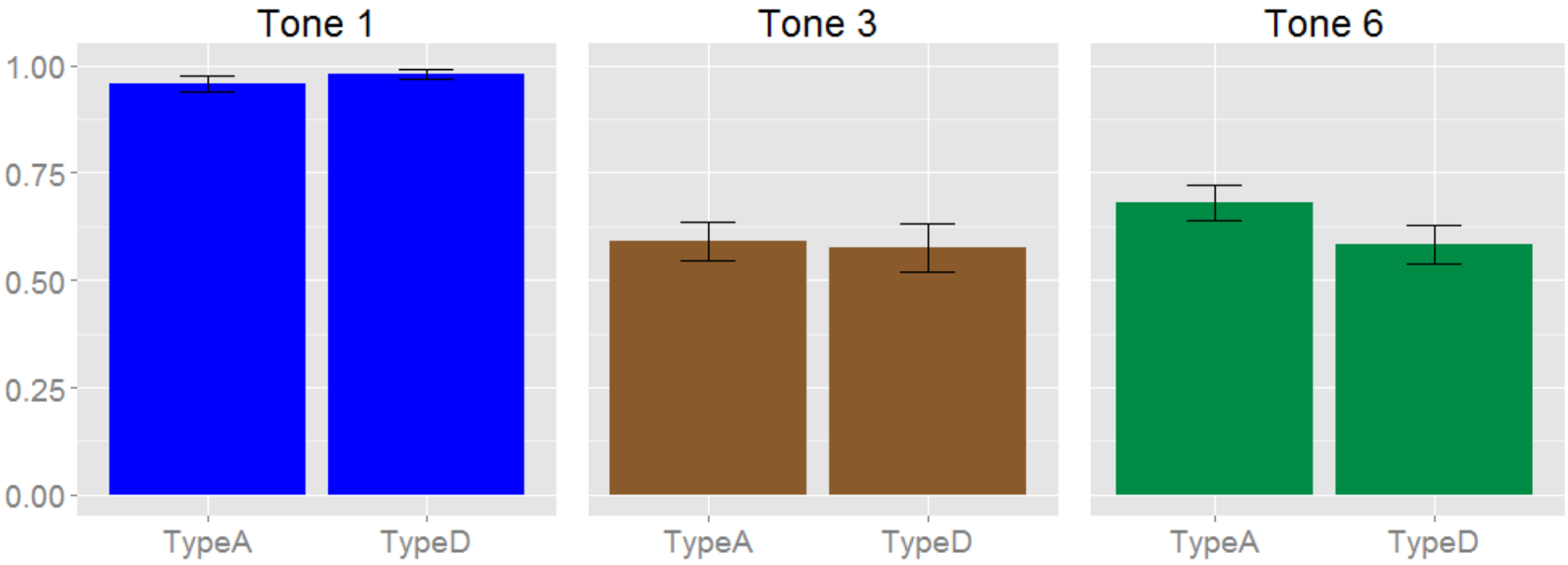
Prediction

Type A > Type D



Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

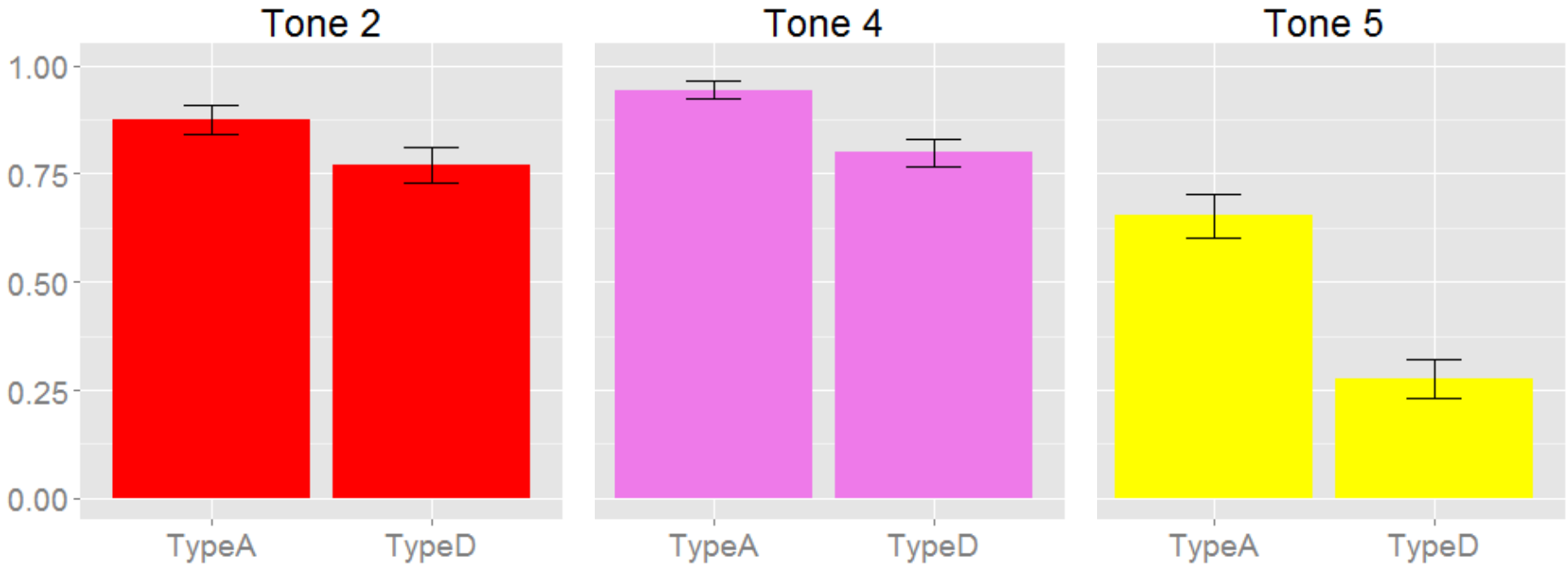
Level tones: Type A (control) vs Type D (2nd half removed)



Prediction: Type A > Type D ✗

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Contour tones: Type A (control) vs Type D (2nd half removed)

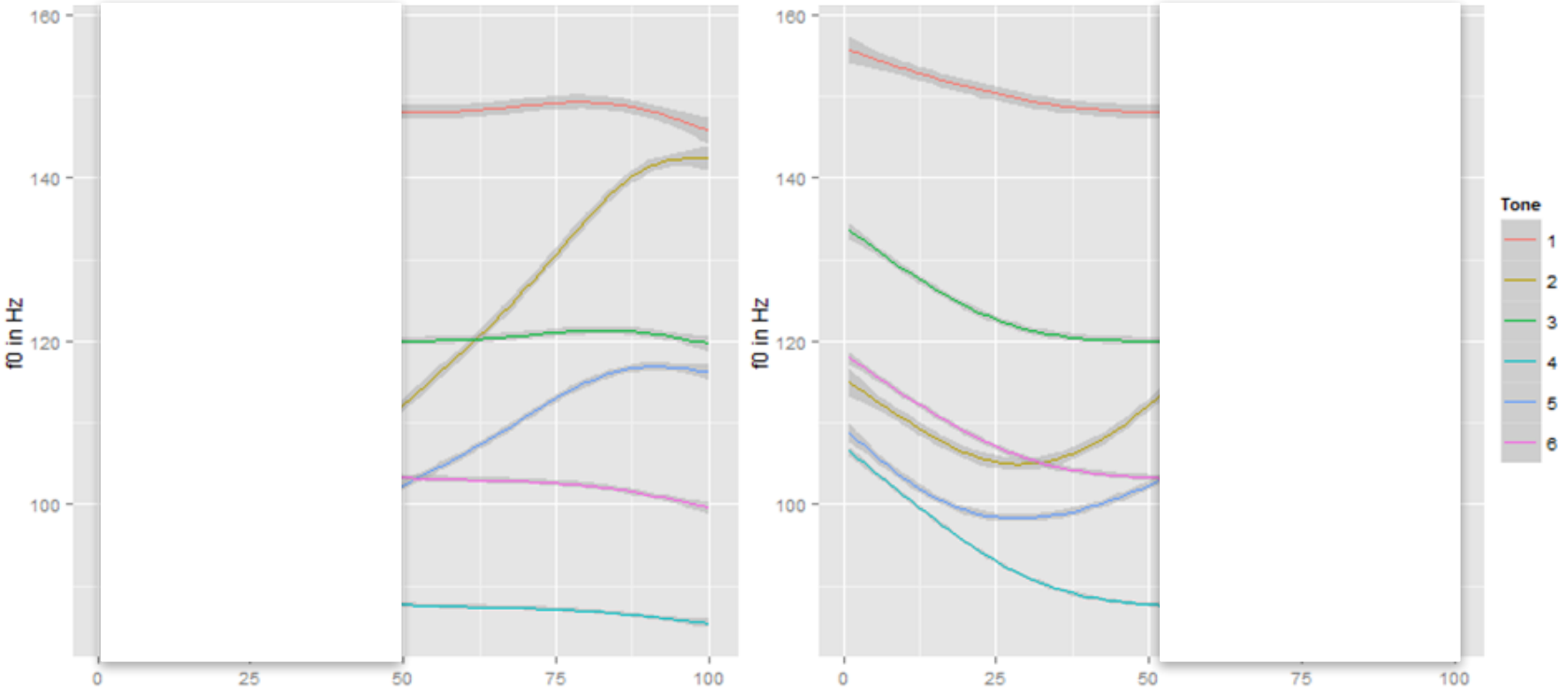


Prediction: Type A > Type D ✔

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

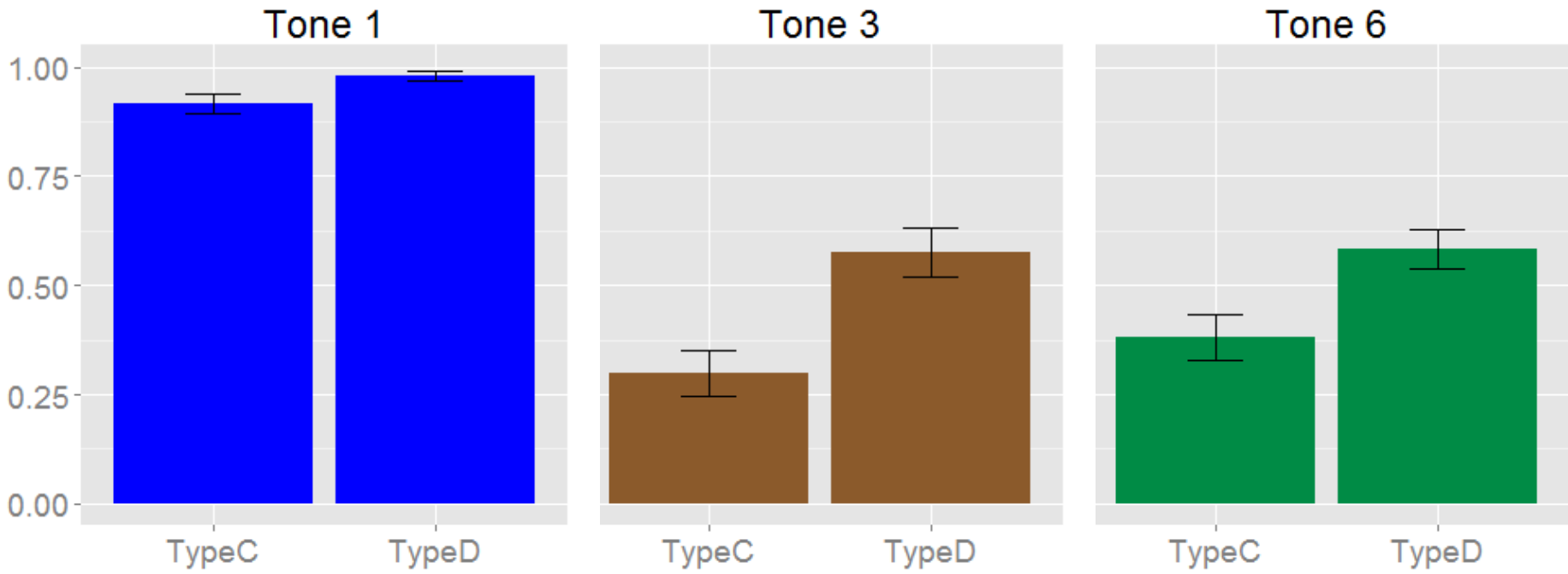
Prediction

Type C > Type D



Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

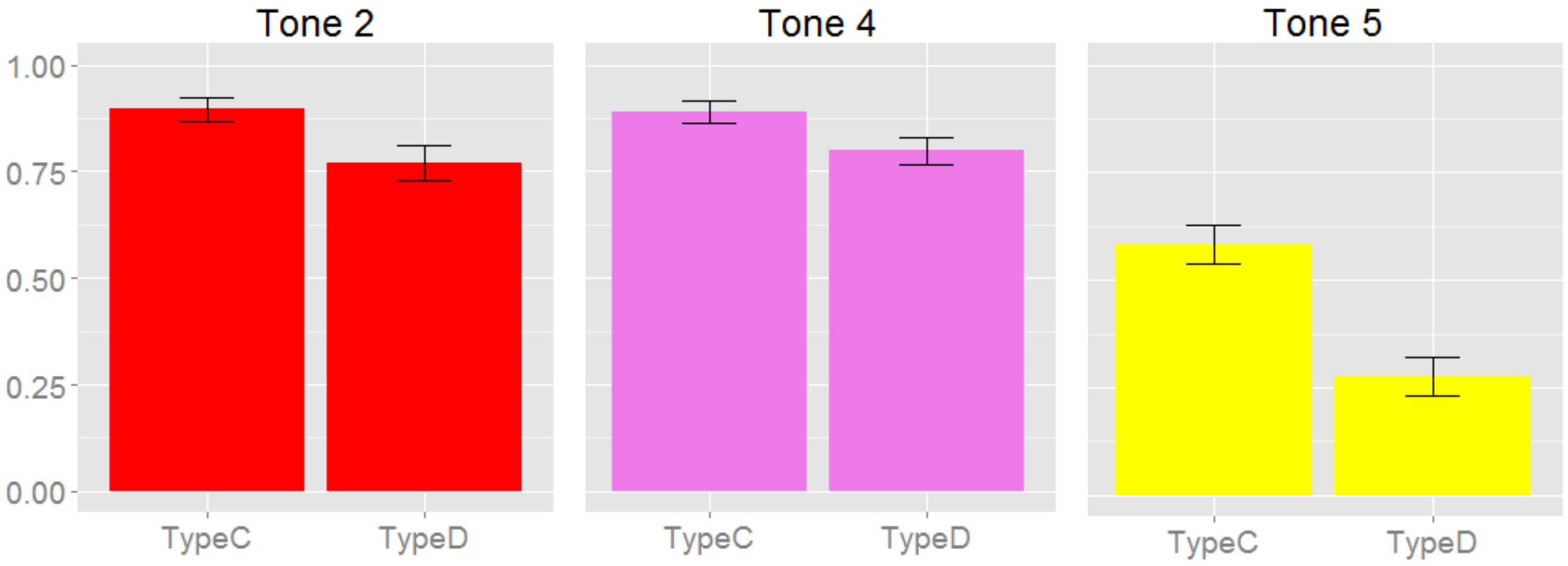
Level tones: Type C (1st half removed) vs Type D (2nd half removed)



Prediction: Type C > Type D ✗

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Contour tones: Type C (1st half removed) vs Type D (2nd half removed)



Prediction: Type C > Type D ✔

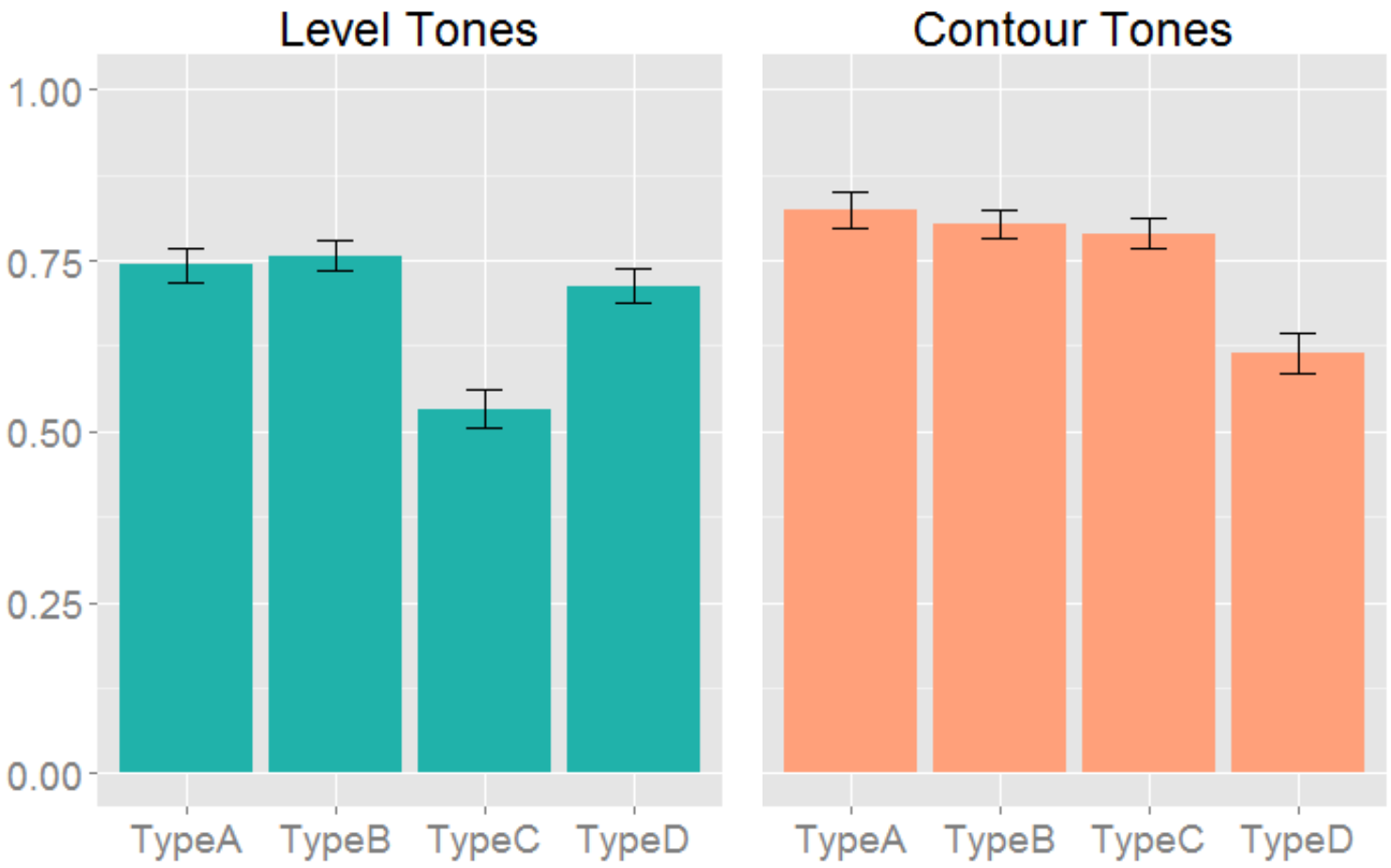
Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Data analysis

- A **logistic mixed effects model** with percentage correct as the dependent variable was fit with fixed effects for **Stimulus Type** (A, B, C, D), **Tone Type** (Level, Contour) and their **interactions**.

- The random effect structure was as maximally specified as possible with random effects for Subject and Talker, and by-Subject random slopes for the interaction of Stimulus Type and Tone Type.

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

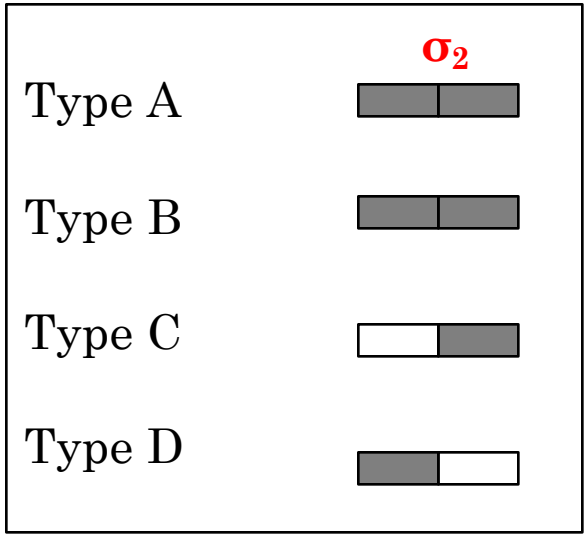


Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Discussion

Listener's bias

When hearing Type C → tend to choose a contour tone
When hearing Type D → tend to choose a level tone

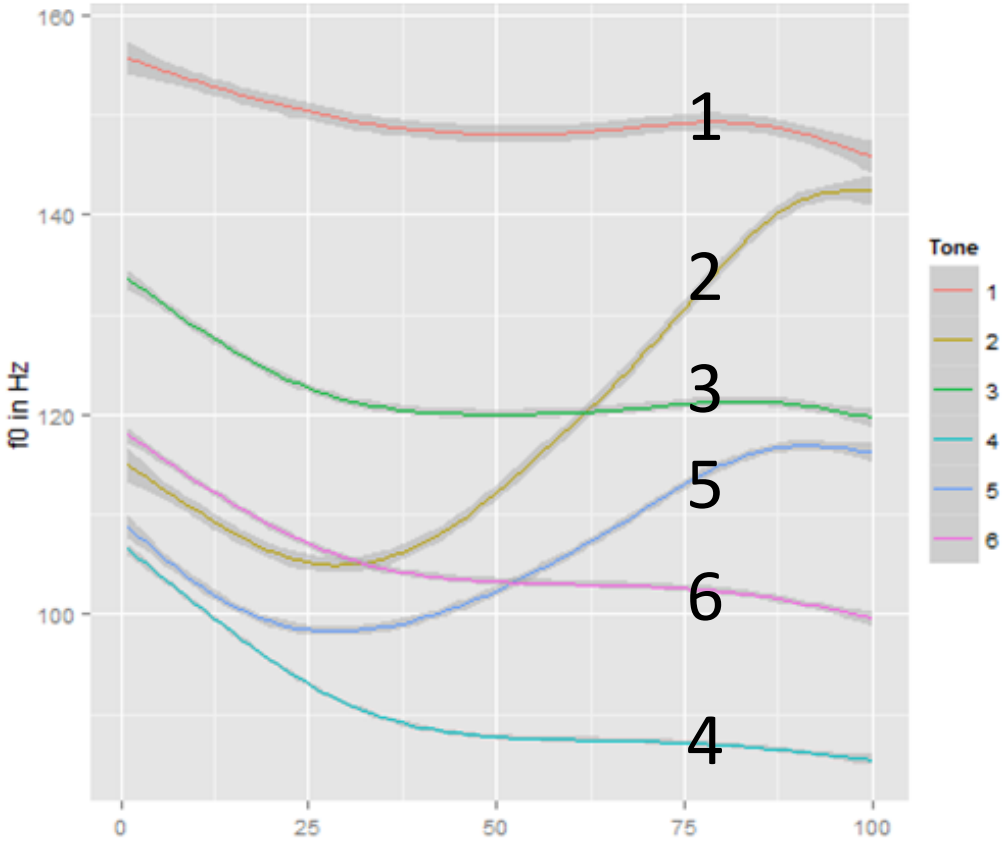


	% of listeners' response	
Type	level	contour
A	50.3	49.6
B	48.3	51.7
C	**37.9**	**62.0**
D	**53.6**	**46.4**

$p < 0.001$
$p = 0.03$

Discussion

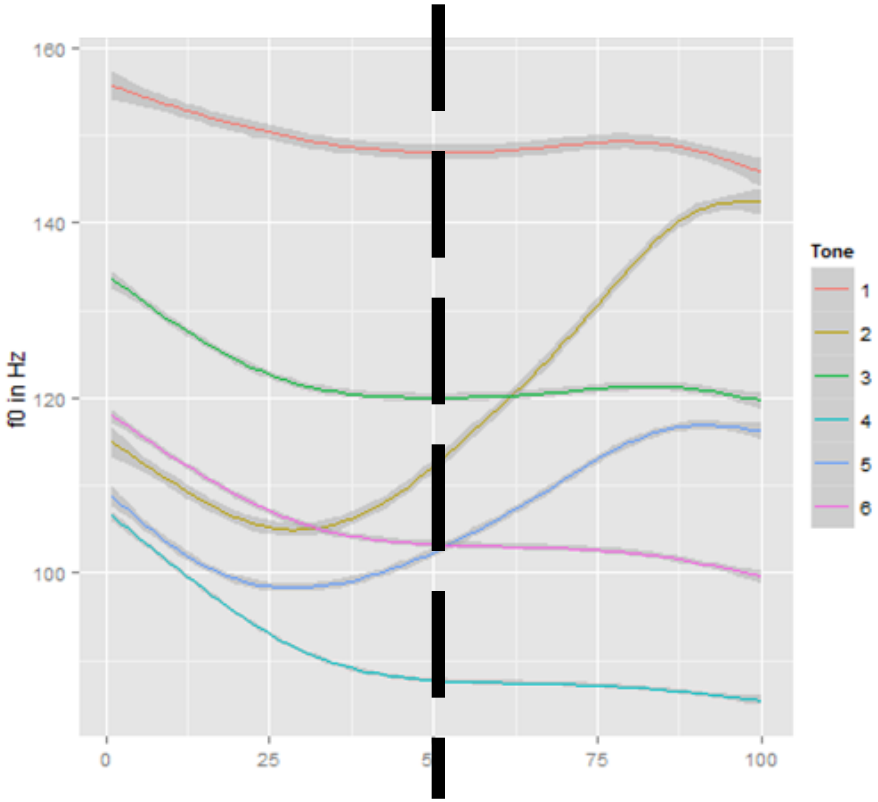
Tone 4 patterns with other contour tones despite being phonetically similar to level tones.



Conclusion

Listeners rely on perceptual cues from different parts of the syllable, **depending on the tone type**.

This half is needed for the identification of **level** tones.



This half is needed for the identification of **contour** tones.

Acknowledgements

This project was supported by a research grant from Social Sciences and Humanities Research Council of Canada (F10-04735).

We would like to thank Molly Babel and participants of the research seminar "Attention and Salience" at the University of British Columbia in 2014-2015, who offered invaluable comments throughout the process. All errors are ours.