Cross-language study of voicing perception of rate-varied bilabial stops

Kyoko Nagao[a] and Kenneth de Jong[b]
[a] *Department of Linguistics and Department of SPHS*
[b] *Department of Linguistics*
Indiana University, Bloomington

At Ling Lunch on 2003 11 18

Acknowledgements

Byung-jin Lim (SUNY)
Minoru Fukuda (Miyazaki Municipal University)
Jin-young Tak (Sejong University)

National Institutes of Health
National Science Foundation

OUTLINE

- Review of previous studies
  - Rate effect on production
  - Rate effect on perception
- Research questions
- Experiments
  - Production
  - Perception
    - English results (Identification tests, Goodness rating)
    - Japanese and Korean results (Identification test)
- Summary and Conclusions

BACKGROUND

1. **Speaking rate affects acoustic properties of voicing categories.**

   (Miller, Green, & Reeves, 1986; Volaitis & Miller, 1992)

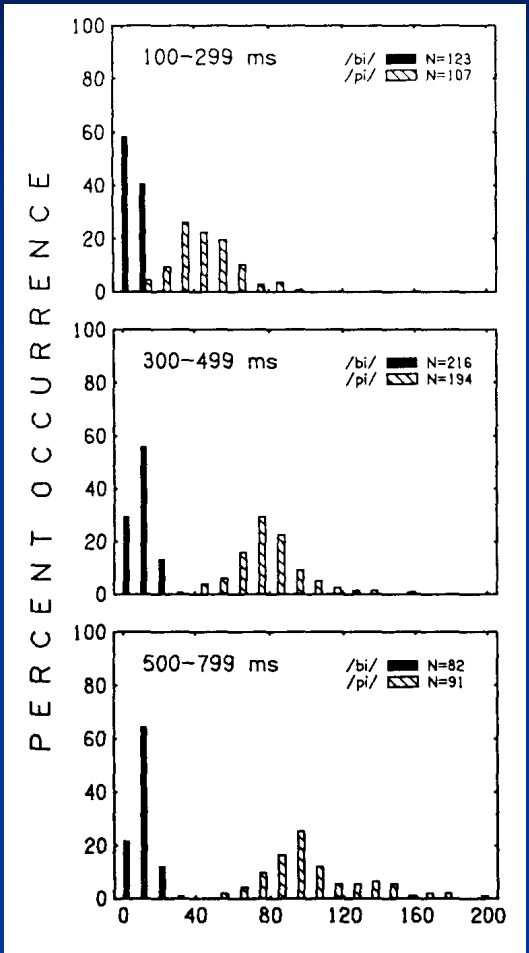
2. **Speaking rate affects perception of voicing contrasts.**
   - **Category boundary shifts.**
   - **Best example shifts.**

   (Miller & Volaitis, 1989; Volaitis & Miller, 1992)

Miller et al. (1986)
Volaitis & Miller (1992)

- A magnitude-production procedure was used to elicit various rates of speech.
  - Normal rate > 2 times fast > 4 times fast > as fast as possible > normal, 2 times slow > 4 times slow > as slow as possible
  - 6 repetitions of each syllable for each rate.
  - Self rate-controlled speech

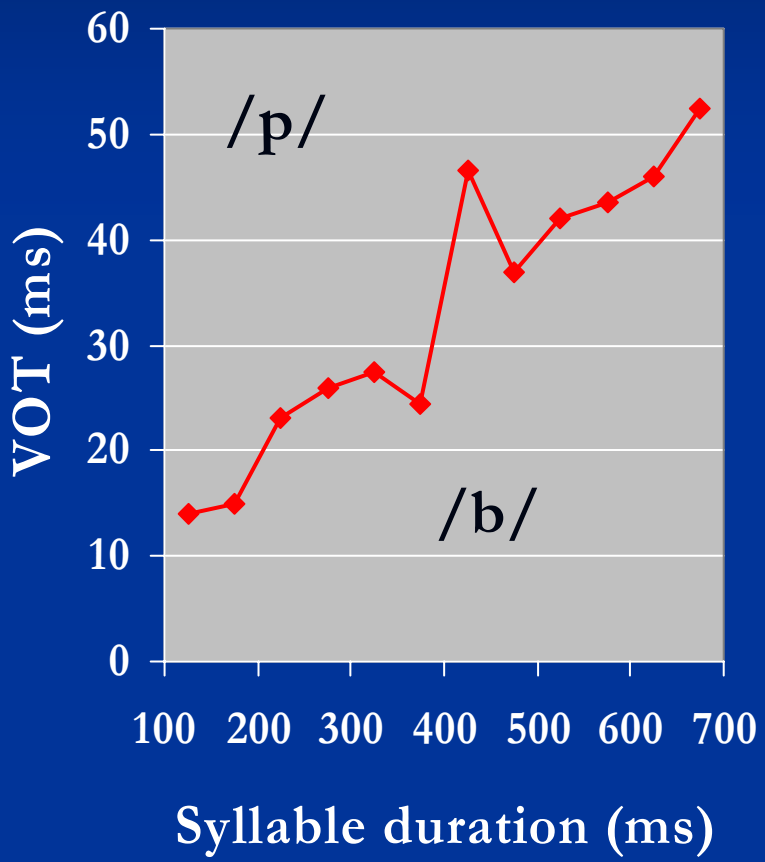
Rate effect on VOT



[Taken from V & M (1992). Percentage of syllables with VOT values within successive 10-ms intervals for three syllable duration intervals. X-axis is VOT.]

- VOT systematically increases with an increase in syllable duration.
- VOT for voiceless consonants change more with rate than voiced consonants.
- For the voiceless stops, the range of VOT distributions increases as syllable duration increases.

Estimated VOT boundary



- Miller et al. (1989) computed VOT values optimally differentiating voicing categories.
- More than 90% of tokens were successfully categorized with rate-dependent optimal VOT values.

Rate effect on perception

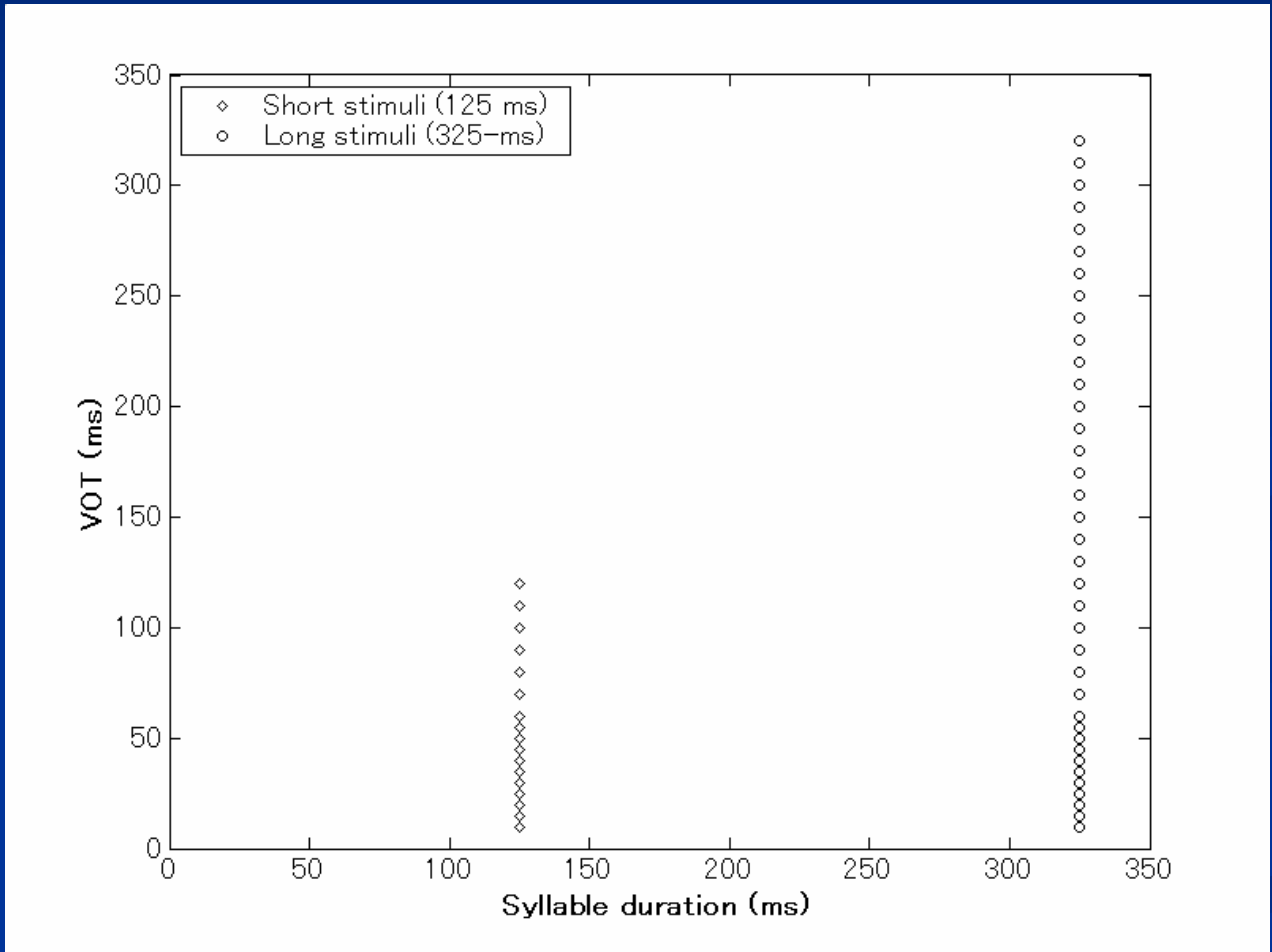
Speaking rate affects perception of voicing contrasts

(Miller & Volaitis, 1989; Volaitis & Miller, 1992)

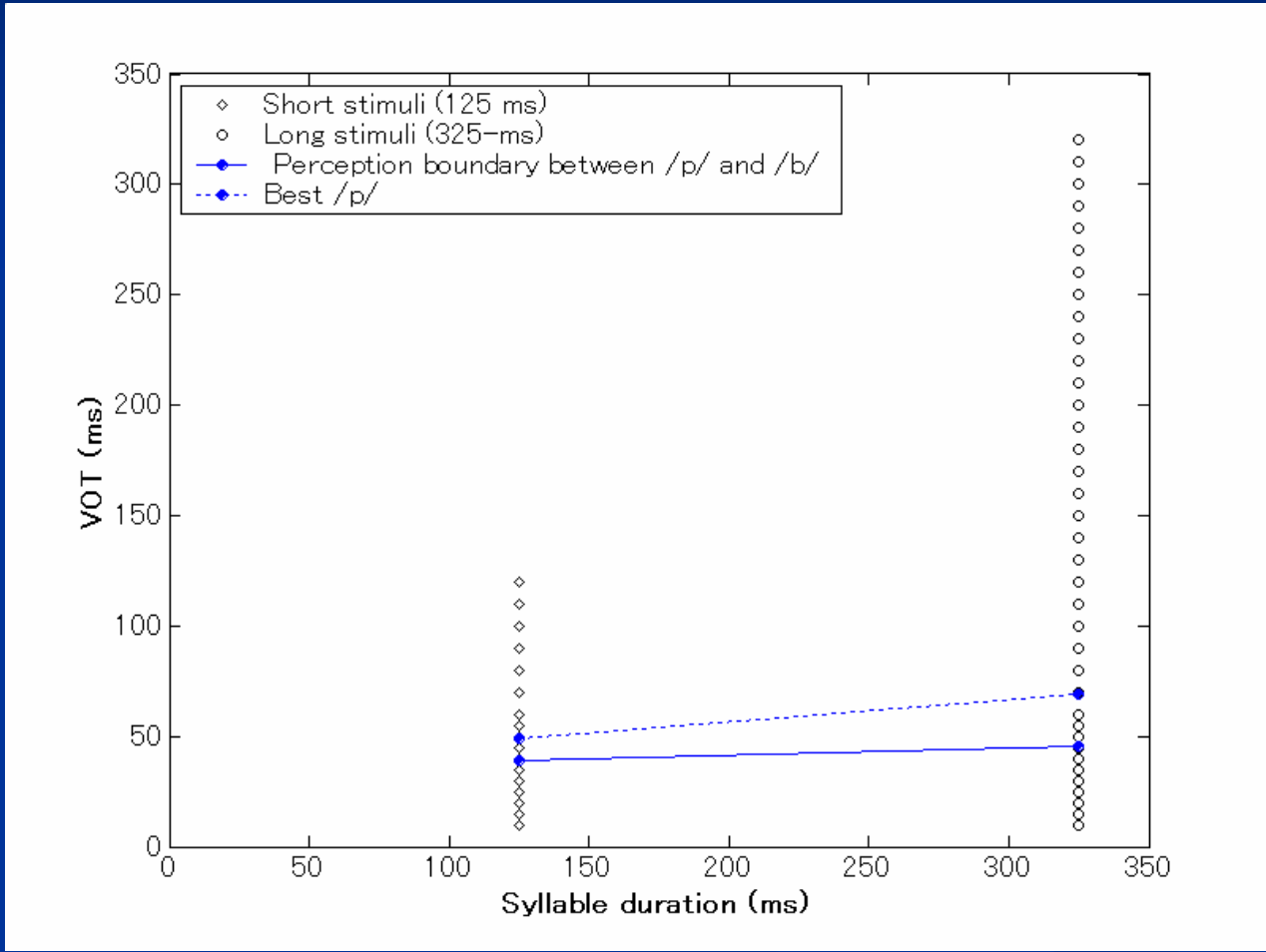
Volaitis & Miller (1992)

- Synthesized syllables with different VOT values.
- Three-forced choice identification test (/p/, /b/, and exaggerated /p/ (called */p/)
- Goodness judgment of /p/ with 10-point scale

Stimuli space of V & M (1992)



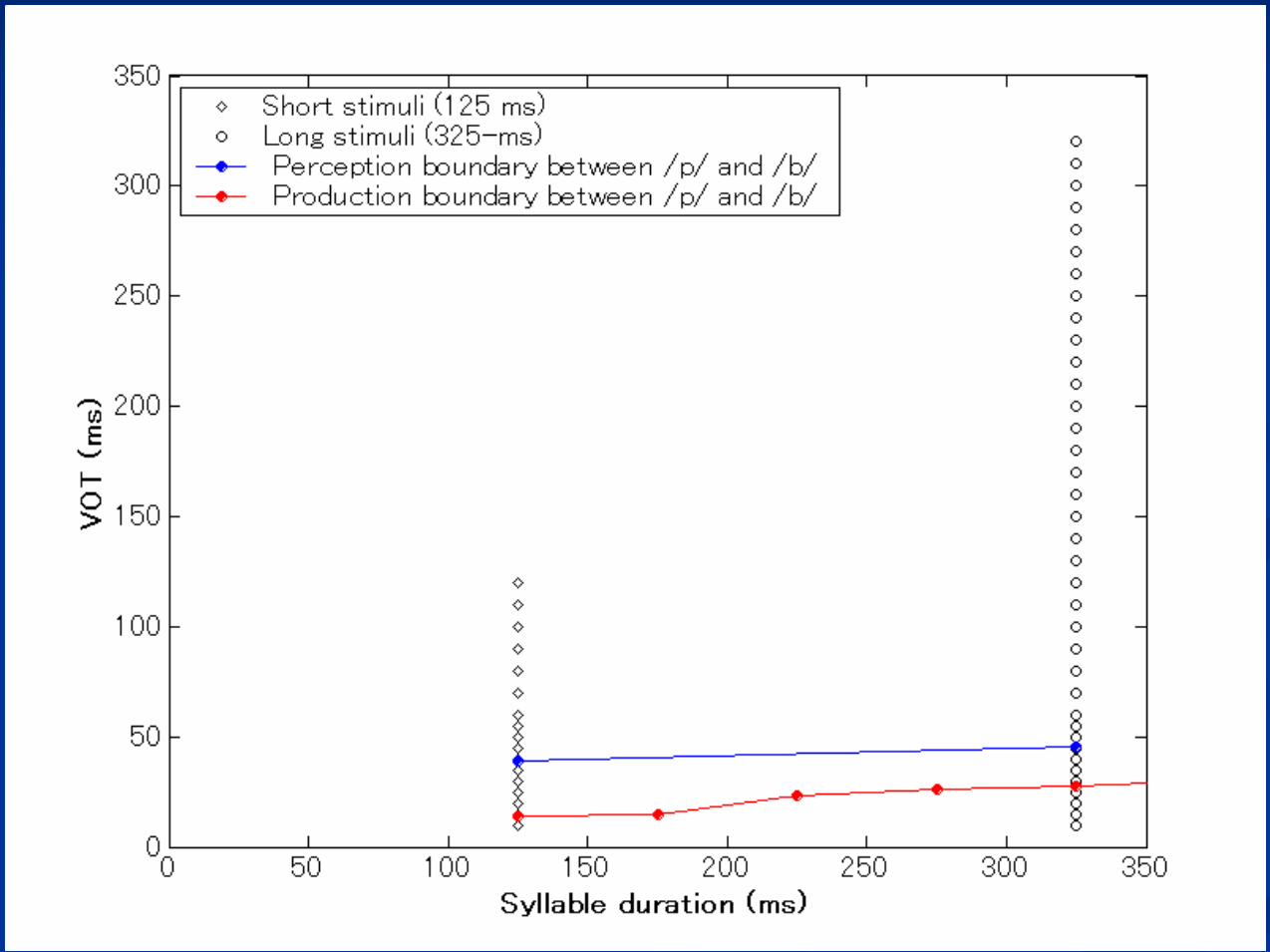
Perceptual boundary (V & M, 1992)



Rate effects on perception

- Changing rate shifts perceptual boundary. The boundary between perceived /b/ and /p/is at a longer VOT when syllable duration is longer.
- Changing rate shifts best example. VOT values of the highest rated /p/ are longer when syllable duration is longer.

Production & Perception results



Problems

- Miller and her colleagues collect various rates of speech by a magnitude-production procedure, but rates are self-controlled.
- Large portion of their speech corpus are long syllable duration.
- Perception results are based on synthesized speech.
- The perceptual boundaries do not match with the VOT values found in literature (Lisker & Abramson, 1970), nor their production results.

RESEARCH QUESTIONS

- Is rate-controlled speech similar to rate-self-controlled speech ?
- Since production and perception studies find different boundaries, how does perception of naturally rate-varied speech compare with studies of previous synthetic speech?
- How do produced rate variation in voicing and perceived rate-normalized voicing judgments correlate with each other?
- Does the same effect happen with listeners with different linguistic backgrounds?

Experiment 1: Method

Speakers: 4 native speakers of Am. English

Recording:

Speakers repeated /bi/ & /pi/ with increasing rate.

Rates were controlled by a metronome.

Measurements:

VOT and Syllable duration were measured.

The fastest 21 syllables of each /bi/ and /pi/ utterance.

Analysis:

Logistic regression analysis with VOT and SYLLABLE DUR

Experiment 1: Results

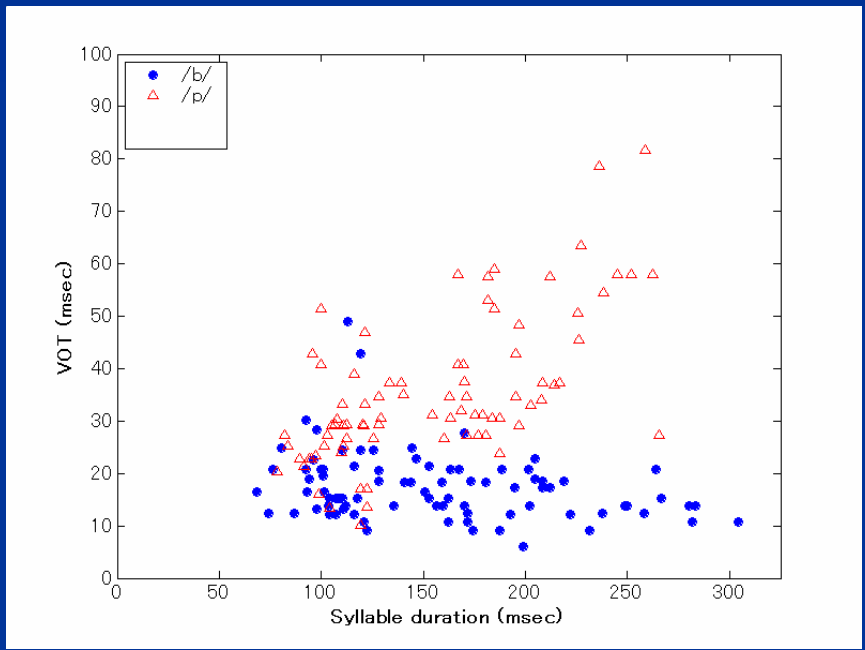


FIG. Distribution of VOT.

- Rate-induced method successfully elicit fast rate of speech.
- As syllable duration increases, VOT values for /p/ increase.
- This rate effect on VOT is larger for /p/ than for /b/.
- Distribution of VOT values of /p/ is wider than /b/.
- VOT values for /b/ and /p/ are overlapped at fast rates.

Experiment 1: Results (Cont.)

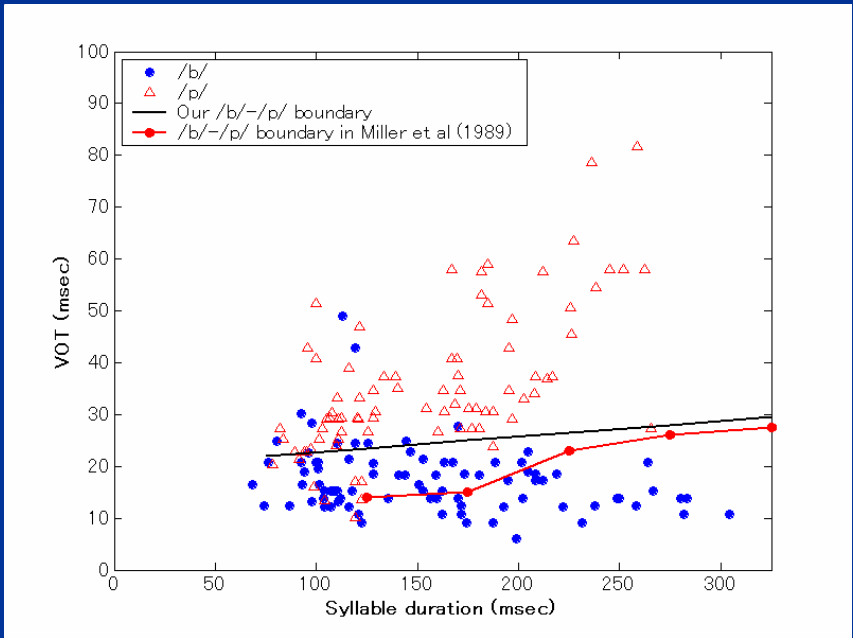


FIG. Estimated boundaries

- Rate-dependent optimal VOT values for /b/-/p/ boundary in Miller et al (1989) does not differentiate categories successfully at fast rates.
- Stimulus VOT ranges in the previous perceptual studies greatly exceeded the ranges in natural speech.

Experiment 2 (ID test): Method

Listeners: 18 native speakers of Am. English

Stimuli:

- 21 stimuli were spliced from each repetitive utterance.
- Each stimulus consists of three repeated syllables.
- VOT and Syllable duration for each stimulus were based on the middle syllable in the stimulus.

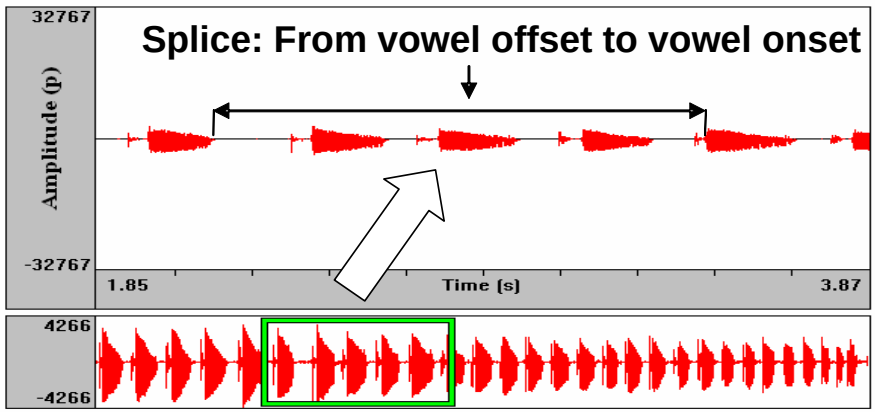


FIG. An example of spliced stimuli /pi/

Experiment 2: Method (Cont.)

Tasks: 4 forced-choice identification test

('*bee*', '*pea*', '*eeb*', and '*eep*')

Analysis:

VOT and Syllable duration of the middle syllable were used for a logistic regression analysis.

Experiment 2: Results

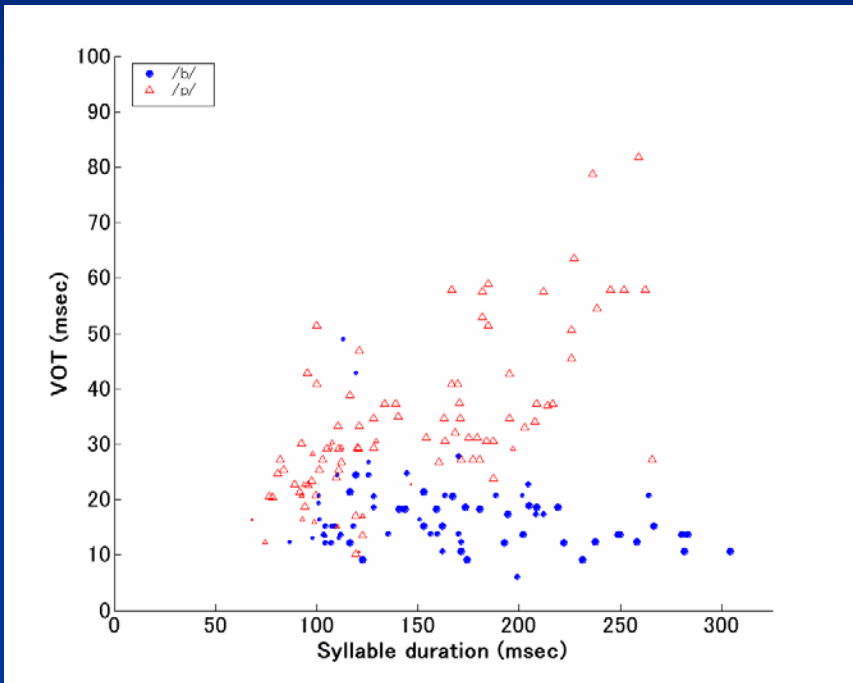


FIG. Relationship b/w VOT and syllable duration based on perception results.

- Identification of /p/ is accurate even at fast rates.
- Accuracy of /b/ ID decreases at fast rates.

Experiment 2: Results (Cont.)

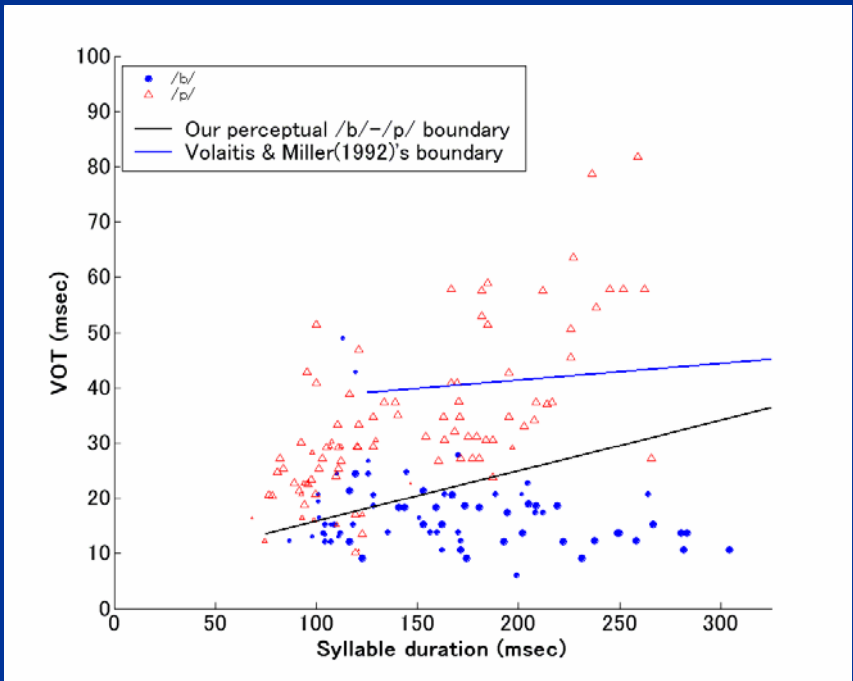


FIG. Estimated perceptual boundaries

- The perceptual boundary matches with VOT boundaries found in EXP 1.
- Positive slope of the estimated boundary indicates that as syllable duration increases, boundary VOT values increase.
- The perceptual boundary from responses to natural speech matches speakers' intended production more accurately than the one found from synthesized speech.

Experiment 3 (Goodness rating)

Listeners: 17 native speakers of Am. English

Stimuli: The same stimuli from EXP 2 were used.

Tasks:

1.Two-alternative forced choice tests, one for consonant identification (*'p' or 'b'*) and one for syllable structure identification (*'Consonant before Vowel'* or *'Vowel before Consonant'*).

2. Goodness rating of identified consonant from 1(=Terrible) to 10(=Excellent).

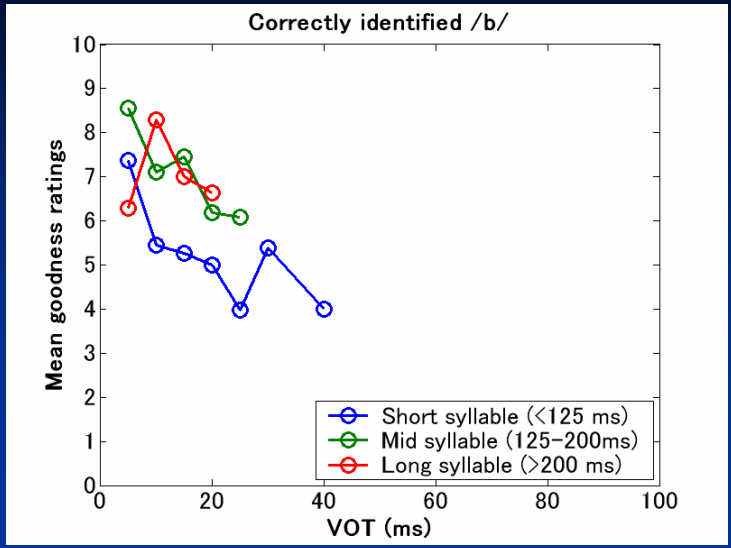
Analysis: Mean goodness ratings were separately calculated for correct and incorrect identification. Stimuli were grouped in terms of values of VOT and syllable duration.

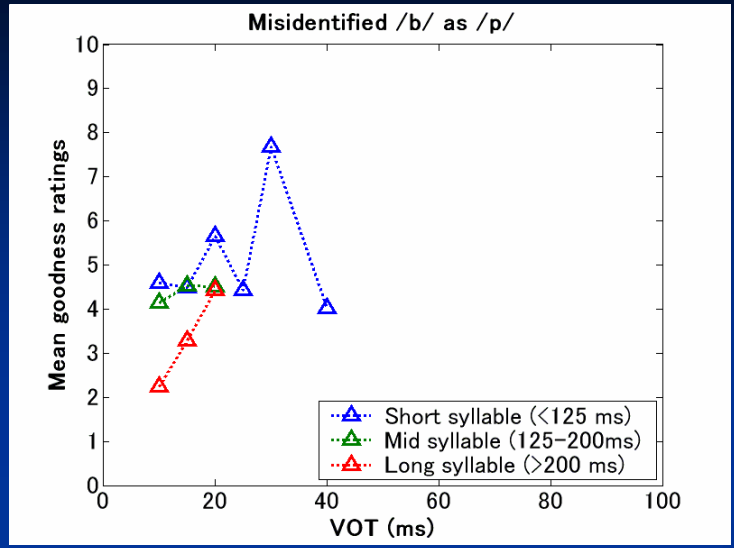
Matched ID

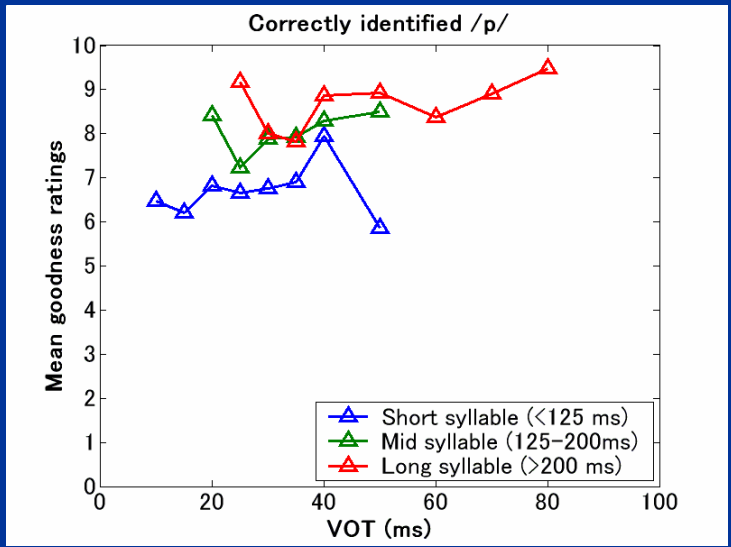
Miss-matched ID

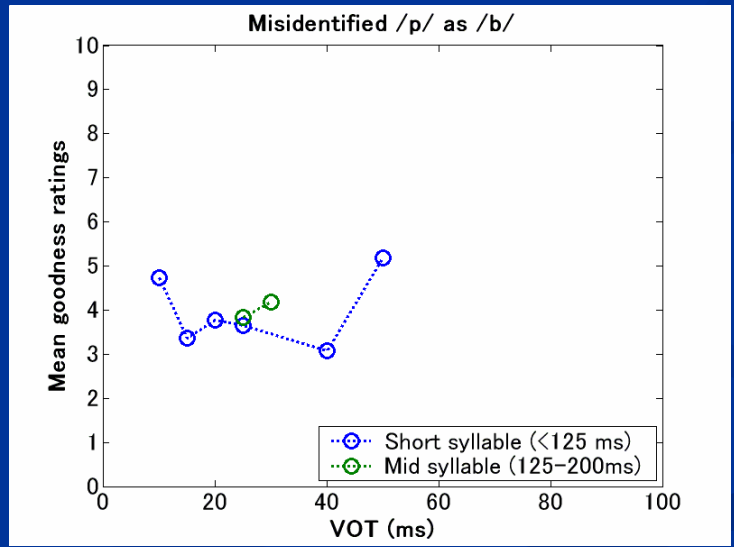
/b/

/p/









O—O: Short syllables O—O: Mid syllables O—O: Long syllables

Conclusion from Exp of English

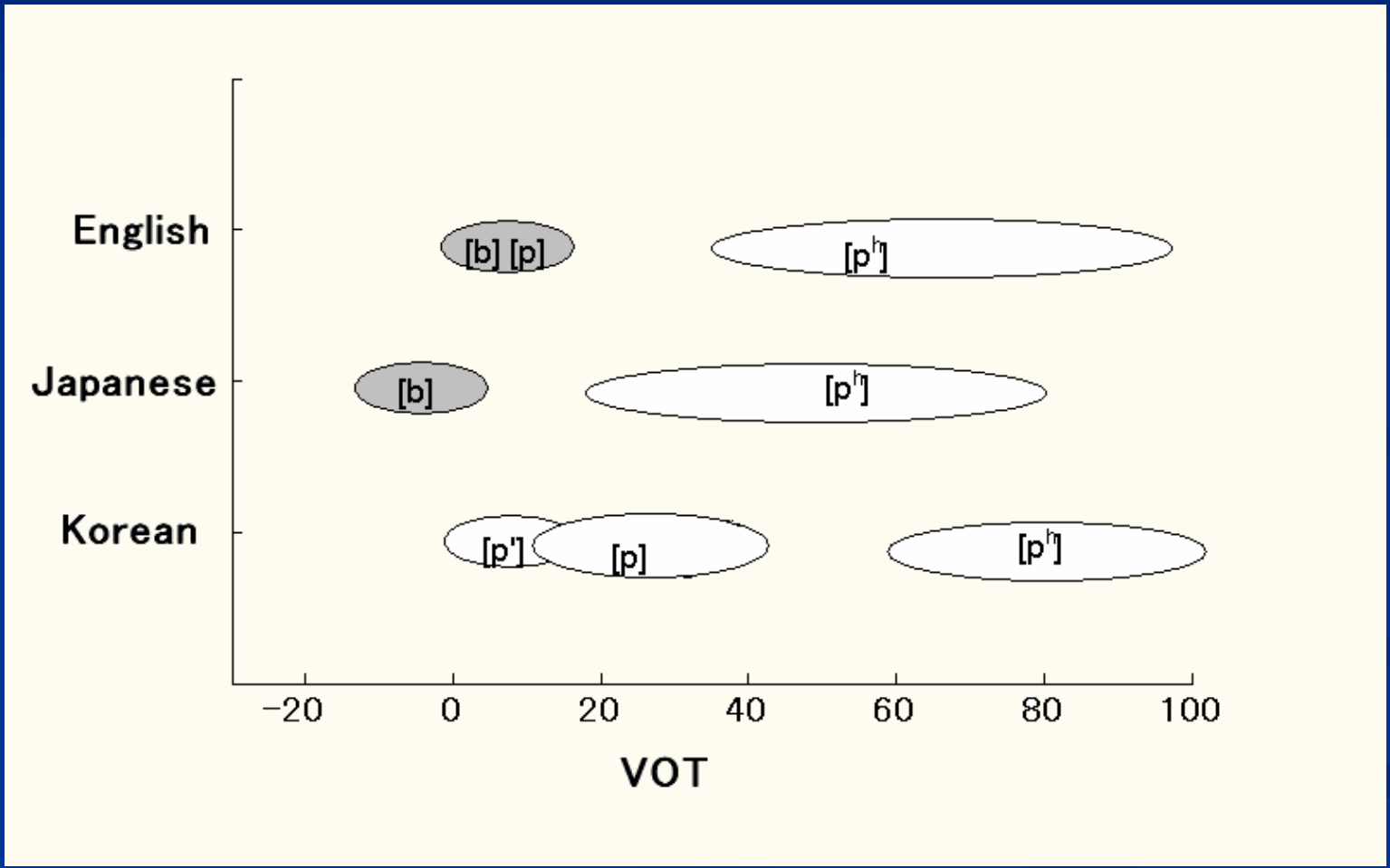
- Rate normalization effects from synthesized speech also occur in the perception of natural speech.
- Speech rate affects both production and perception in a similar manner. The perceptual identification system is neatly tuned to the distributions found in production.
- The results of goodness ratings indicate that listeners store fine-grained information to distinguish voicing contrasts.
- Accurate identification of segments with aberrant VOT values suggests listeners use signal attributes in addition to VOT to differentiate the contrast.
- Even though rate-varied repetitive speech is uncommonly encountered, listeners effectively deal with rate-induced variation in categorization tasks. This suggests an active component in listeners' perceptual systems which generalizes to novel circumstances.

- Is this a part of general auditory mechanism?

  If so,

  We expect the same results from native listeners of other language.

Schematic phonetic categorization for voicing on the VOT continuum



Experiment 4 (Non-native listeners)

Listeners:

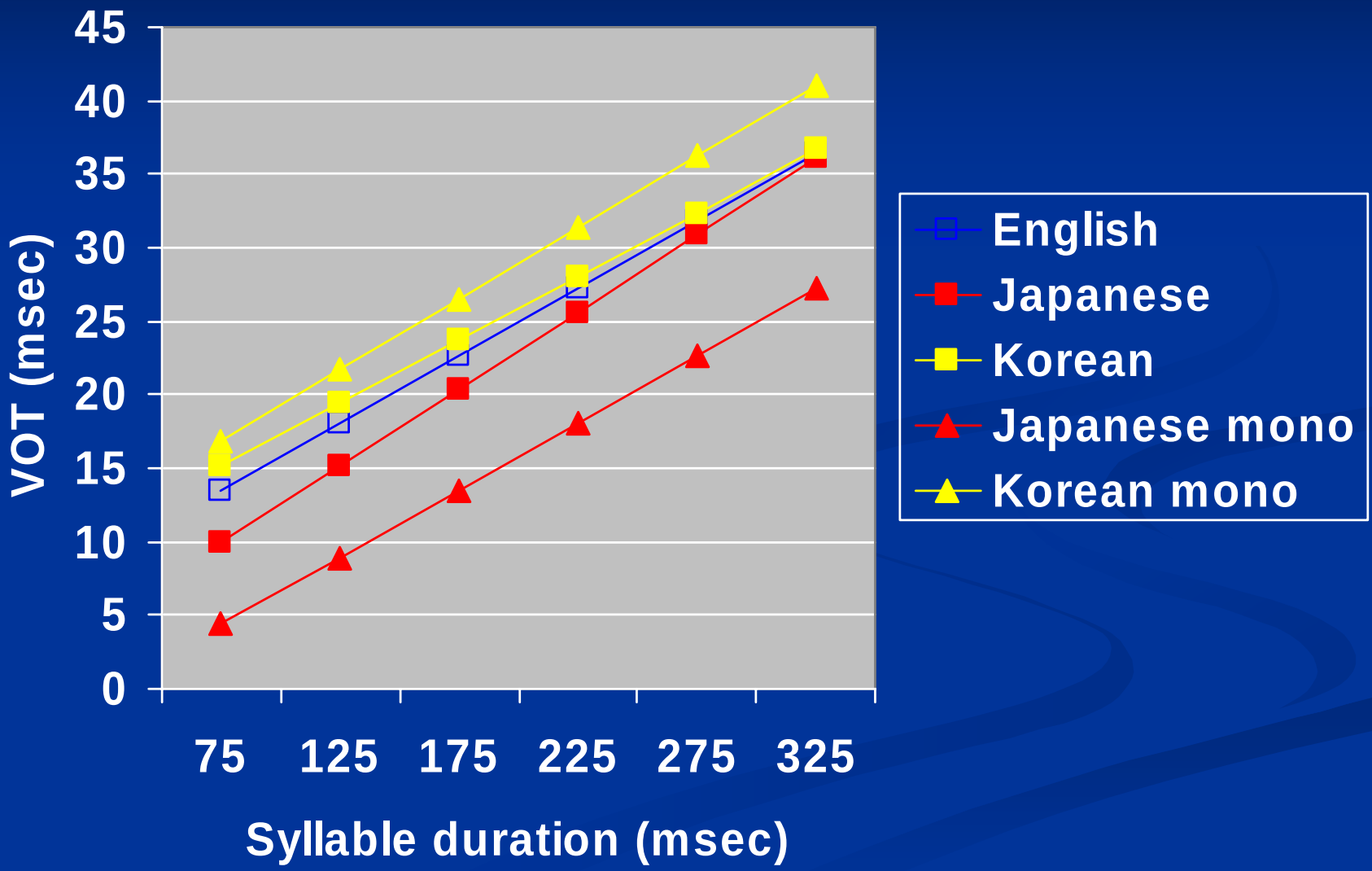
- 34 native speakers of Japanese
  - Bilingual Japanese: 14 were recruited in the US.
  - Monolingual Japanese: 20 were recruited in Japan.
- 33 native speakers of Korean
  - Bilingual Korean: 14 were recruited in the US.
  - Monolingual Korean: 20 were recruited in Korea.

Stimuli: The same stimuli from EXP 2 were used.

Tasks: 4 forced-choice ID test: (‘*bee*’, ‘*pea*’, ‘*eeb*’, and ‘*eep*’)

Analysis: VOT and Syllable duration of the middle syllable were used for logistic regression analysis.

Estimated /b/-/p/ boundaries



Conclusion

- Rate normalization is not a general auditory mechanism. It is based on the distribution of consonants that the listeners have experience.

Implication from X-lang results

- Learners of English could have a difficulty to perceive voicing distinction which is though as relatively easy distinction.
- Even advanced learners might have misperceived voicing categories more than native listeners do when speech becomes fast.

Estimated /b/-/p/ boundaries