A43. Learning Variation in a Second Language:
A Cross-language Study of Rate-normalization

Kyoko Nagao†, Kenneth de Jong†, and Byung-jin Lim‡
†Indiana University
‡University of Wisconsin

Introduction

Specific Issues

- How do non-native perceivers cope with variability in second language categories?
  - Non-native listeners have been reported to have particularly large intelligibility deficits under non-optimal (non-laboratory) listening conditions.
- What is the nature of phonological categories?
- How do they get modified in the face of experience with a new language?

Types of Variation in Speech Categories

- **Bio-physical variations:**

  Variation due to the nature of the articulatory system (e.g. acoustic details of formant transitions, etc). All experienced listeners have access to this regardless of native language.

- **Linguistic variation:**

  Variation due to language (e.g., phonologically specified allophonic alternations, etc). Listeners must acquire the ability to account for these when learning a second language.
  - For example, VOT at the boundary to distinguish voicing categories differs across languages (e.g. English vs. Japanese vs. Korean).
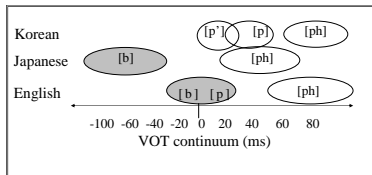

FIG 1. Schematized figures showing differences in VOT values for Korean, Japanese, and English stops.

Speech Rate/Tempo

- Distinction of voicing categories is rate-sensitive because VOT, the strongest voicing cue, is influenced by speaking rate.
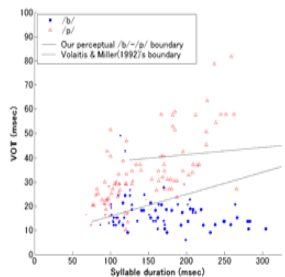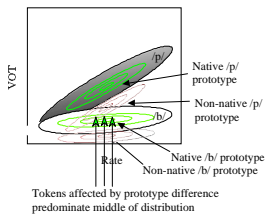- Listeners account for the rate effects to extract the voicing categories.

FIG 2. Distributions of VOT values for various syllable durations for English /b/ (blue) and /p/ (red). The straight line is the estimated perceptual boundary for English listeners. [Reprinted from Nagao & de Jong, 2003.]

Acknowledgements

Research Questions

- How do second language learners respond to highly rate-varied tokens in judging non-native voicing categories?
- How is this effect modulated by increased experience with the second language?
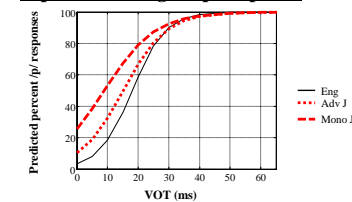
Hypothesis

- If rate-induced variations are accounted for by a general auditory mechanism, we would expect no deviations from native perceptions in non-native perceptions.
- If normalization is based on the distribution of consonants that the listeners have experienced, we expect deviations from native perceptions. This deviation is less if the listeners have experience of consonant distributions in both native and non-native languages.
- If the internal structure of phonological categories is based on prototype matching, deviations should be most clear in less extreme rates.

Native /p/ prototype
Non-native /p/ prototype
Native /b/ prototype
Non-native /b/ prototype
Tokens affected by prototype difference predominate middle of distribution

- If the internal structure of phonological categories is distributional, deviations of non-native responses would be observed across the entire distributions.

L2 Criterion (English)
L1 Criterion (Japanese)
Tokens affected by criterion mismatch throughout distribution

Methods

Listeners

Non-native listeners (L2=English)

Japanese listeners
- **Advanced ESL learners of Japanese (Adv J):** 14 native speakers of Japanese in the US.
- **Monolingual Japanese (Mono J):** 20 native speakers of Japanese in Japan.

Korean listeners
- **Advanced ESL learners of Korean (Adv K):** 14 native speakers of Korean in the US.
- **Monolingual Korean (Mono K):** 20 native speakers of Korean in Korea.

Native listeners (Control)
- 18 speakers of American English

Listeners	Age (years old)	Duration stayed in the US (months)
Mono. Jap.	18-60 (M=32.5)	Almost none
Adv. Jap.	21-31 (M=24.4)	3-84 (M=40)
Mono. Korean	19-25 (M=23)	Almost none
Adv. Korean	24-31 (M=28)	3-72 (M=40)
English native	18-23 (M=20)	-

Stimuli

Speakers
- 4 Speakers of American English
- From the Northern Midwest
- All in late 20's

Speech corpus
- Speakers repeated /bi/ and /pi/ with increasing rate.
- Rates were controlled by a metronome.
- 21 stimuli were spliced from each repetitive utterance.
- Each stimulus consists of three repeated syllables.

Measurements
- VOT and Syllable duration were measured.
- VOT and Syllable duration for each stimulus were based on the middle syllable in the stimulus.

Procedure

- Task: 4 forced-choice identification test ('bee', 'pea', 'eeb', and 'eep') ('bi', 'pi', 'ib', and 'ip' for Japanese)

Analysis:

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

Results

1. Boundary shift

Japanese vs. English perceptions
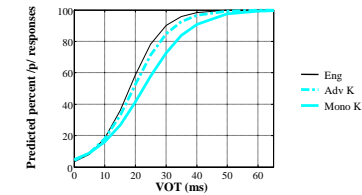

Korean vs. English perceptions

FIG 3. Percent /p/ perception of native listeners predicted by a logistical regression model as a function for VOT continua with a syllable duration of 125 msec.

1. Category boundary was shifted toward boundaries in listeners' native language for both Japanese and Korean listeners.
2. The shift is more apparent for the monolingual listener groups than the advanced ESL learner groups.
3. Both monolingual groups showed less steep boundaries.
4. Learners of English could have a difficulty perceiving the voicing distinction, even though previous studies find little errors with clear lab-style speech.

2. Rate effects
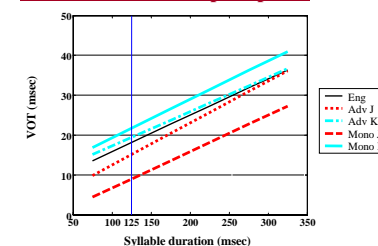### Native vs. Non-native perceptions

FIG 5. VOT values at the estimated category boundary for each listener group as a function for syllable duration (msec).

5. Rate normalization appeared in all groups.

Results (Cont.)

3. Distributional Analysis:
Distributions of Deviations of Adv. ESL from Native Listeners
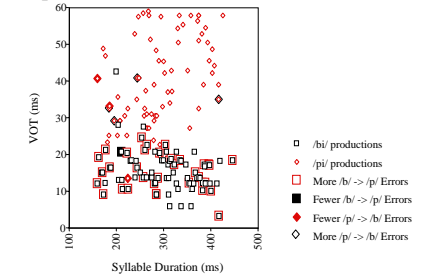
Japanese listeners

FIG 6a. Distributions of VOT values for various syllable durations for /b/ (squares) and /p/ (diamonds). Outlining of the token symbols indicates a 20% difference between English listeners and Japanese listeners.

- Japanese listeners more likely to mislabel /b/ as /p/. (Red tokens flow into /b/ productions at bottom).
- Deviations occur across the entire distribution.
- Some errors in /p/ distribution as well.
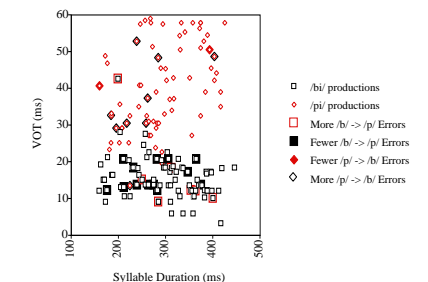
Korean Listeners

FIG 6b. Distributions of VOT values for various syllable durations for /b/ (squares) and /p/ (diamonds). Outlining of the token symbols indicates a 20% difference between English listeners and Korean listeners.

- Koreans are more likely to mislabel /p/ as /b/. (Black tokens in /p/ region at top)
- Surprisingly, some /b/ are more accurately identified by Koreans than by native listeners.
- Deviations occur across the whole distribution.
- Some errors in /b/ distribution as well.

Summary

- Non-native and native listeners show same rate normalization effects.
- Experiment provides evidence that rate normalization is not a general auditory mechanism. It is based on the distribution of consonants that the listeners have experienced.
- Even contrasts believed to be 'easy' could be misperceived when non-natives encounter rate-related variation.
- Distribution of deviations across the entire range of tokens suggests a generalized criterion function, which gets shifted for the second language.
- However, the large number of error increases in the wrong direction (/p/ -> /b/ for Japanese, and /b/ -> /p/ for Korean) indicate another aspect of second language perception: generalized uncertainty.

References

Nagao, K., & de Jong, K. (2003). Perceptual rate normalization in naturally produced bilabial stops, Presented at the ASA meeting in Austin, TX.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. Journal of the Acoustical Society of America, 92(2), 723-735.