Examining the relationship between the interpretation of age and gender across languages

Andrew R. Plummer[1], Benjamin Munson[2], Lucie Ménard[3], Mary E. Beckman[1]

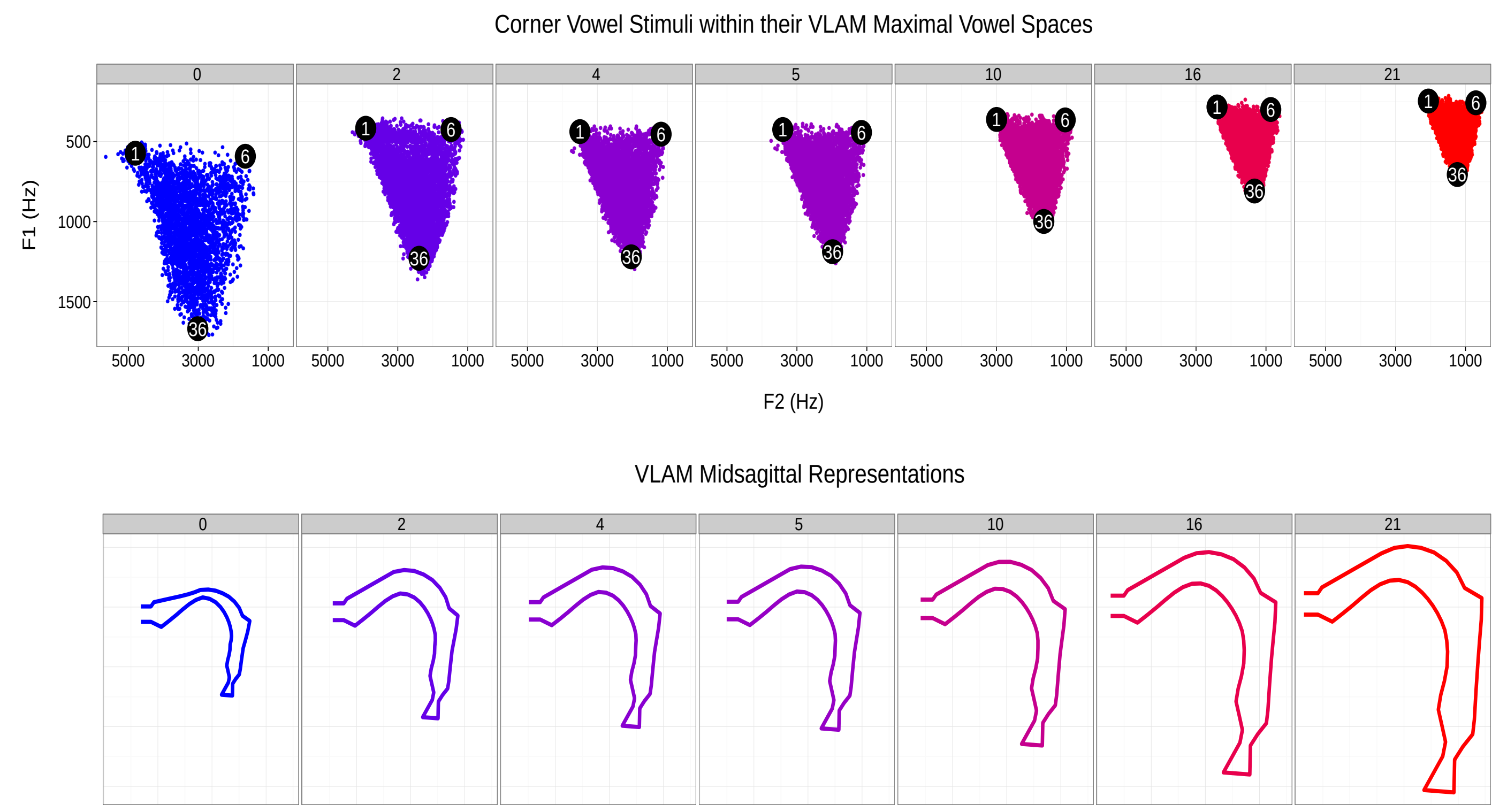
The Ohio State University, Columbus, OH, USA[1],
University of Minnesota, Minneapolis, MN, USA[2],
Université du Québec à Montréal, Montréal, Québec, CA[3]

Introduction

- Speech signals vary substantially in a number of their key properties, with the variability deriving from, among other things, talkers' age and gender.
- Speech processing requires resolution of this variation, necessitating interpretation of age and gender information in the signal.
- In some signals, the age and gender are not clear from acoustic information alone. In these cases there may be substantial individual variation in judgments of age and gender.

In this study, we examine the interplay between the interpretation of age and gender across language communities.

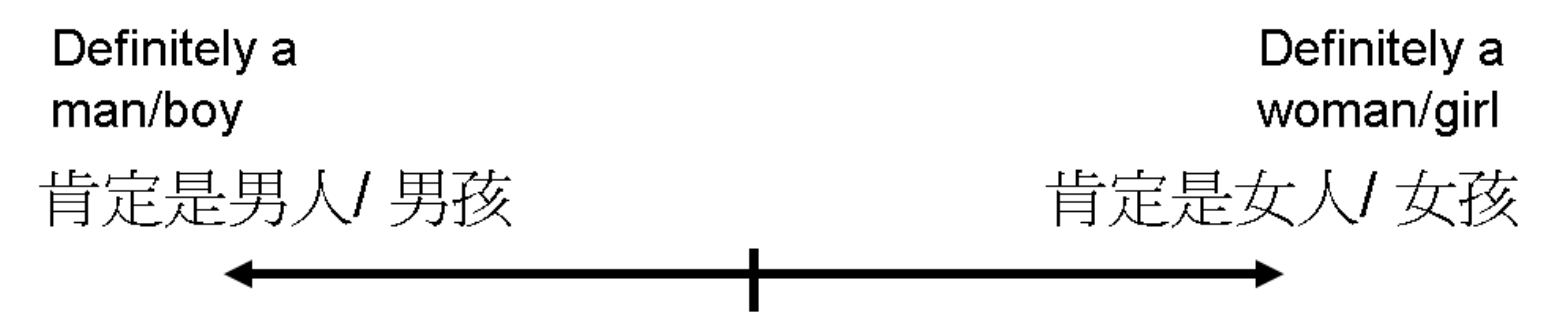
VLAM Corner Vowel Stimuli



- The *Variable Linear Articulatory Model* (VLAM, Boë and Maeda, 1997) is a computational model of the articulatory system and its speech production capacities.
- Midsagittal representations, such as those depicted above, are wrought by configuring "articulatory blocks" (Maeda, 1990; 1991) corresponding to jaw height, tongue body position, tongue dorsum position, tongue apex position, lip protrusion, lip height, and larynx height.
- The VLAM is age-varying and capable of representing articulatory lengths ranging from those of infants to young adults. Given an age in years, the set of all articulatory configurations of the VLAM at that age that do not result in occlusion of the oral cavity yield a corresponding *maximal vowel space* (MVS, Boë et al., 1989) for that age.
- Corner vowel stimuli ([i], [u], [a]) were generated by the VLAM set at seven different ages: 6 months, 2, 4, 5, 10, 16, and 21 years, and are indexed numerically as 1, 6, and 36, respectively, in each of the maximal vowel spaces pictured above.
- Each stimuli set was part of a larger set of 38 vowel prototypes selected from the maximal vowel space for each age.

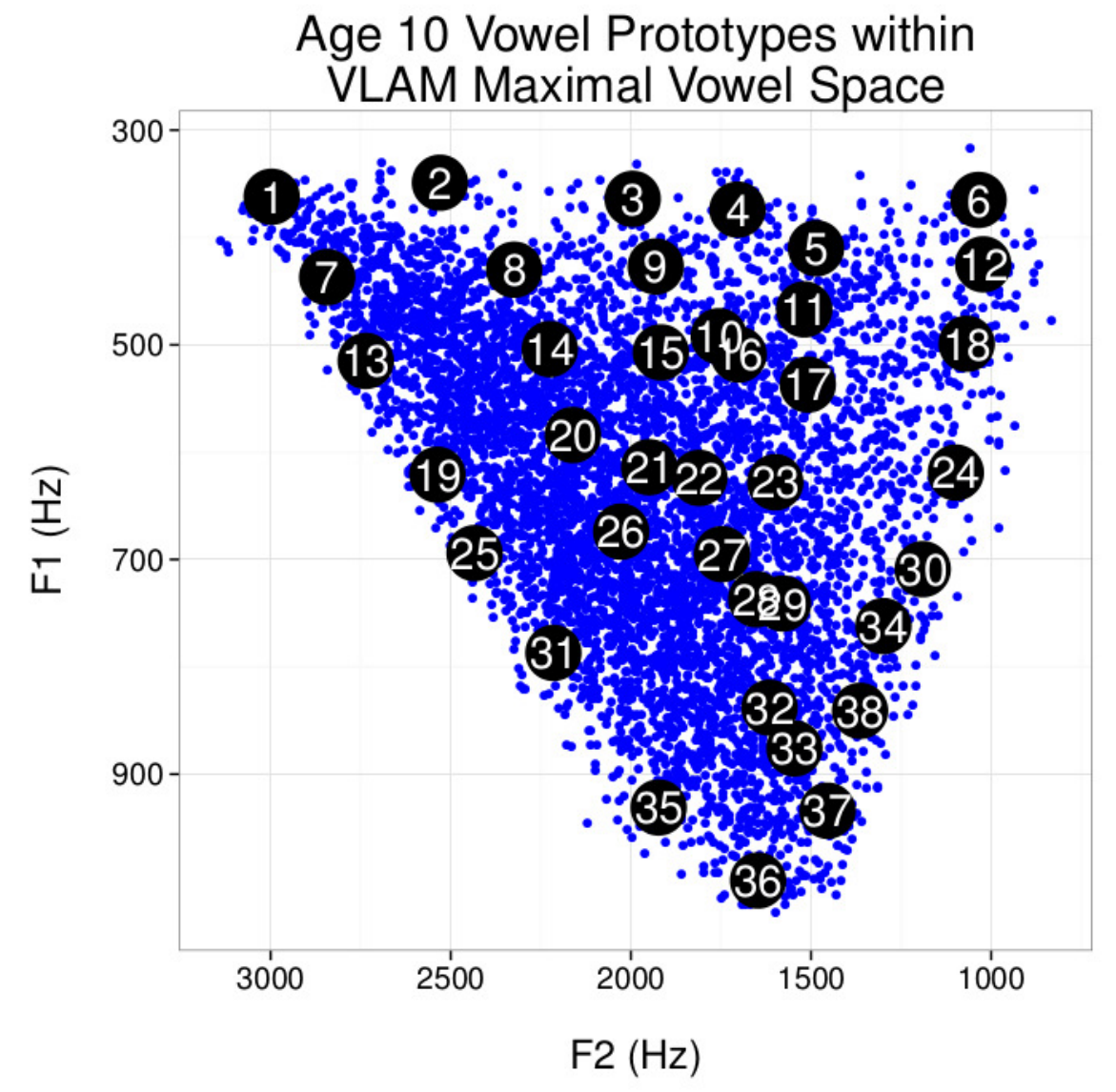
Experimental Procedure

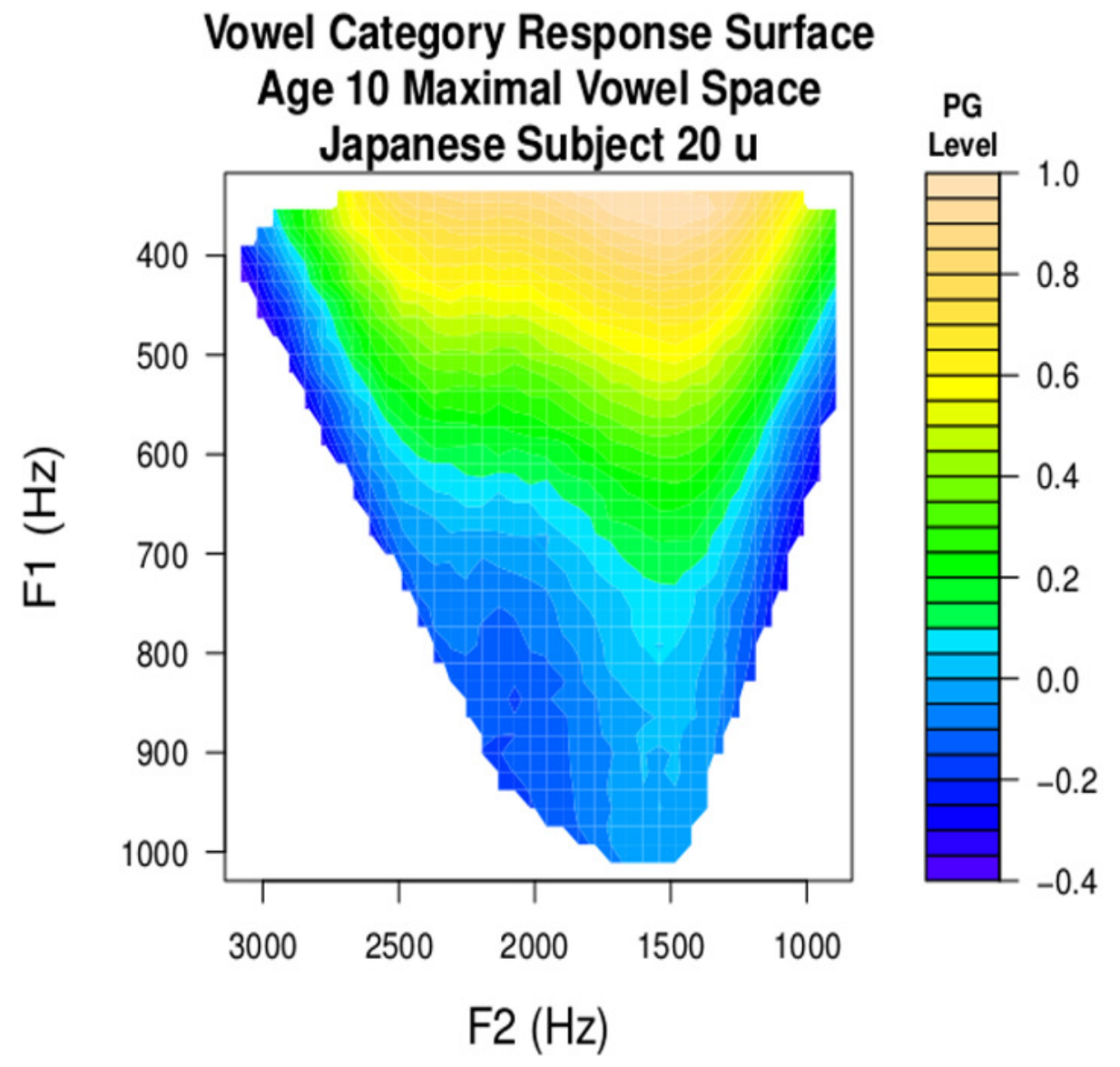
- Subjects included 15 native speakers of Cantonese, 21 native speakers of American English, and 21 native speakers of Japanese.
- Subjects were played each stimulus and each assigned an age in years to each stimulus. Subjects also assigned a gender rating along a visual analog scale ranging from "definitely male" to "definitely female," or the equivalent in Japanese or Cantonese, to each stimulus.



- Responses along the gender scale were converted to numbers ranging from "definitely male" (0 on the scale) to "definitely female" (650 on the scale).

Broader Experimental Paradigm



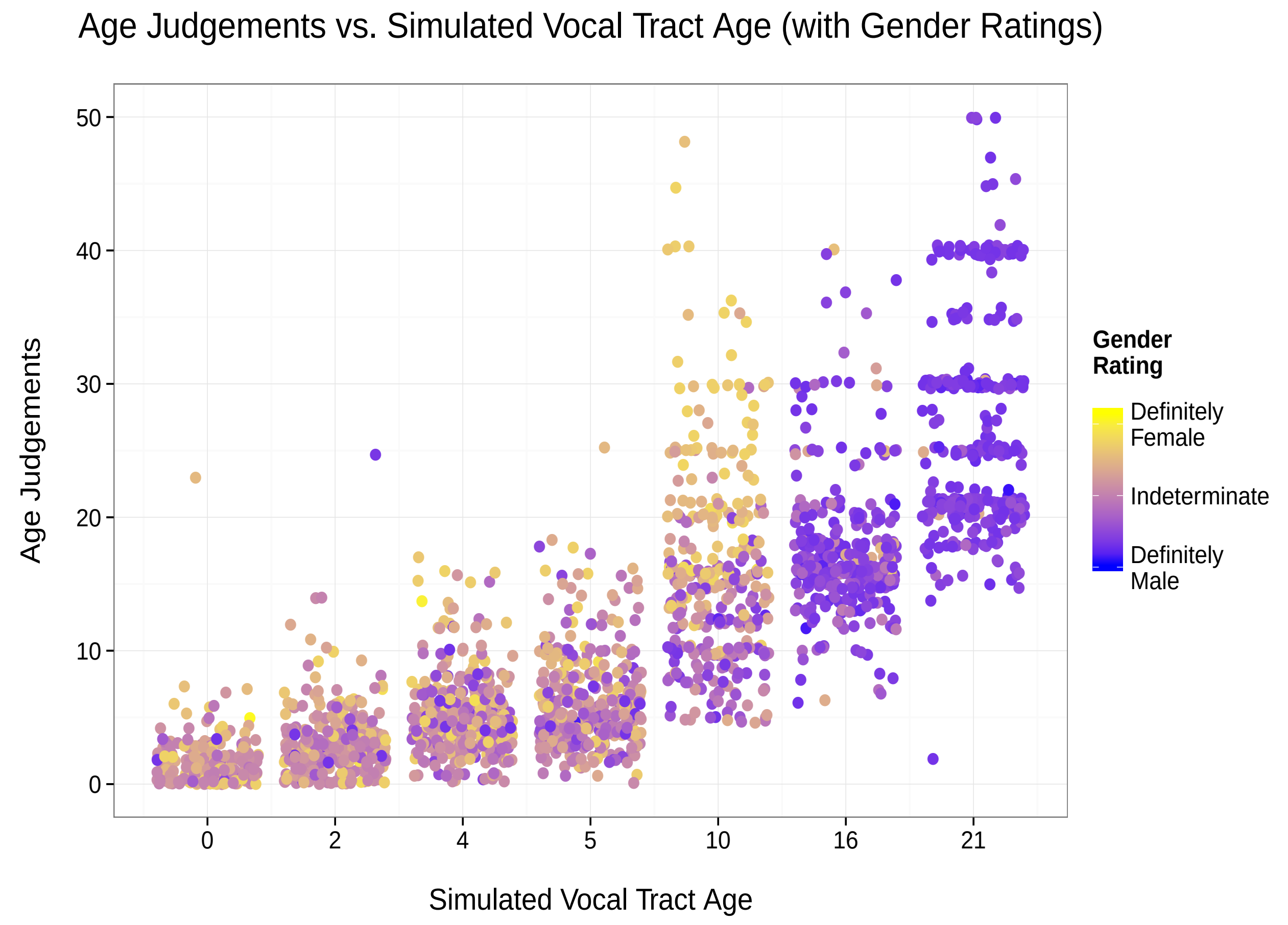


- The age and gender ratings were a final task block in an experiment in which the subjects from each language group first identified each of the larger set of 38 stimuli, blocked by vocal tract age, as one of a set of vowel phoneme categories for that language.
- Prior to each block, subjects were told the vocal tract age of the child whose vowels they would listen to in that block. The order of vocal tract ages was randomized across subjects.
- Listeners responded by clicking on a keyword (for English and Cantonese), or a hiragana symbol unambiguously representing the vowel in isolation (Japanese) (see Munson et al., 2010).
- The cross-cultural differences in age-gender ratings in the current analysis can help illuminate the cross-language differences in the responses to the vowel categorization task for the larger stimulus set, which we have quantified in other work (Plummer, Ménard, Munson, and Beckman, 2013).

Acknowledgements

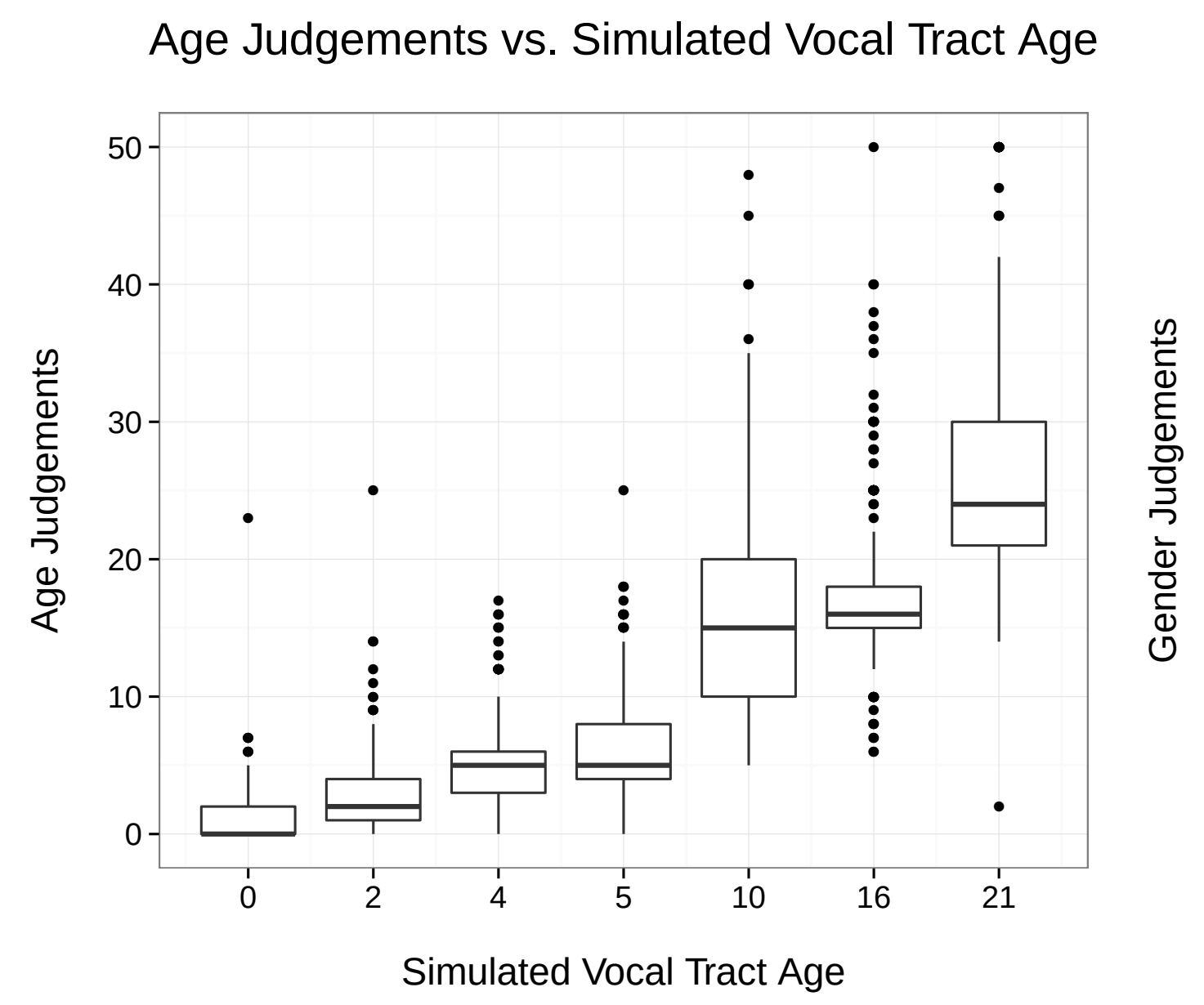
Subjects were recruited in Hong Kong by Benjamin Munson, in Columbus by Chanelle Mays, and in Tokyo by Kiyoko Yoneyama. We thank Catherine McBride and Amy Chan for help recruiting Cantonese subjects. Work supported by NSF grants BCS 0729277 (to Munson) and BCS 0729306 (to Beckman).

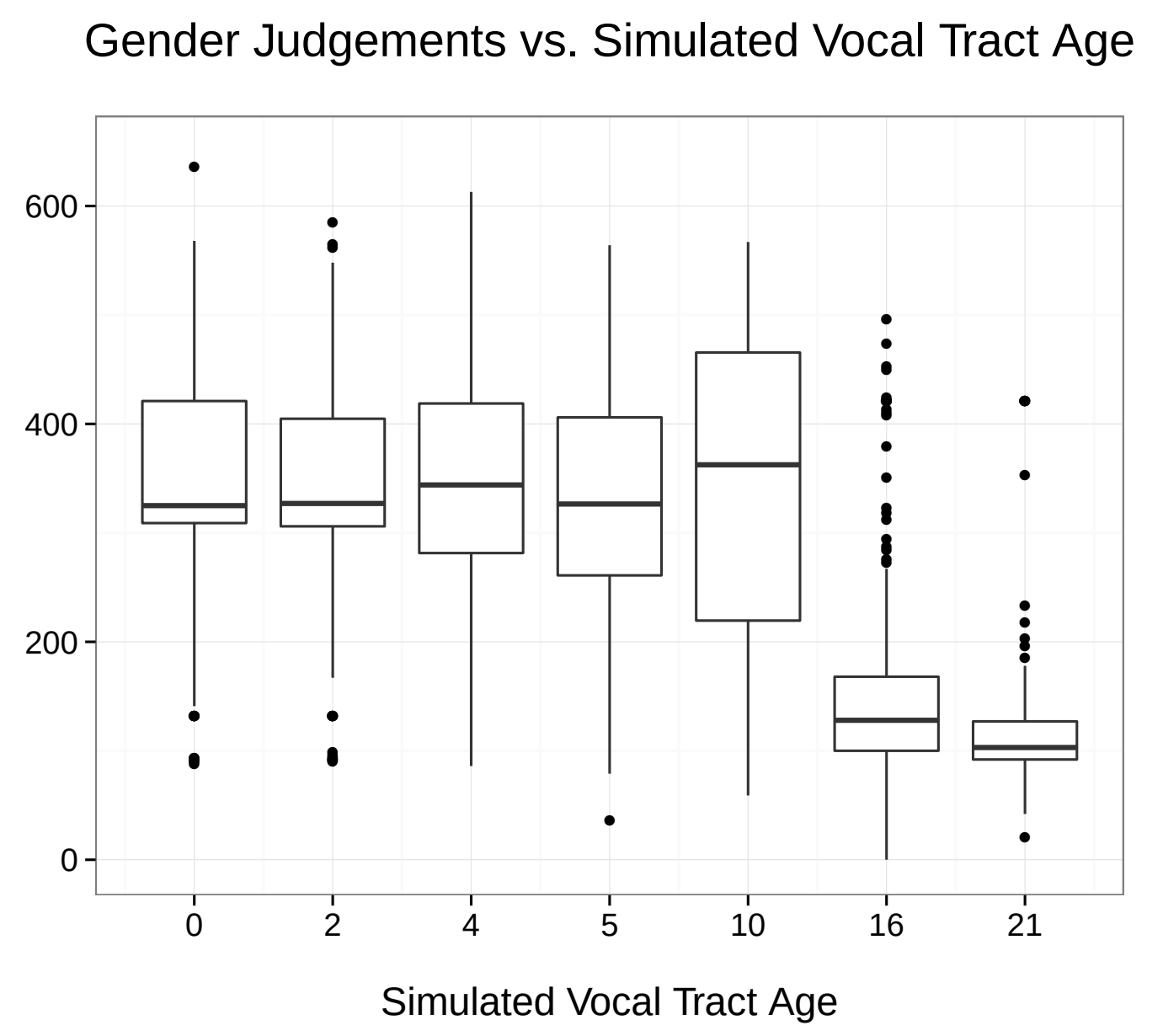
The Interplay between Age and Gender Responses



- A column scatter plot grouped by simulated vocal tract age for the age judgments from all 57 listeners to all 6 trials for each vocal tract age, with the color indicating the assigned gender rating.
- A bifurcation is evident in the interpretation of age and gender for the age 10 stimuli, which subjects rated either as a younger male or an older female, suggesting a nonuniformity in the resolution of variability during processing.

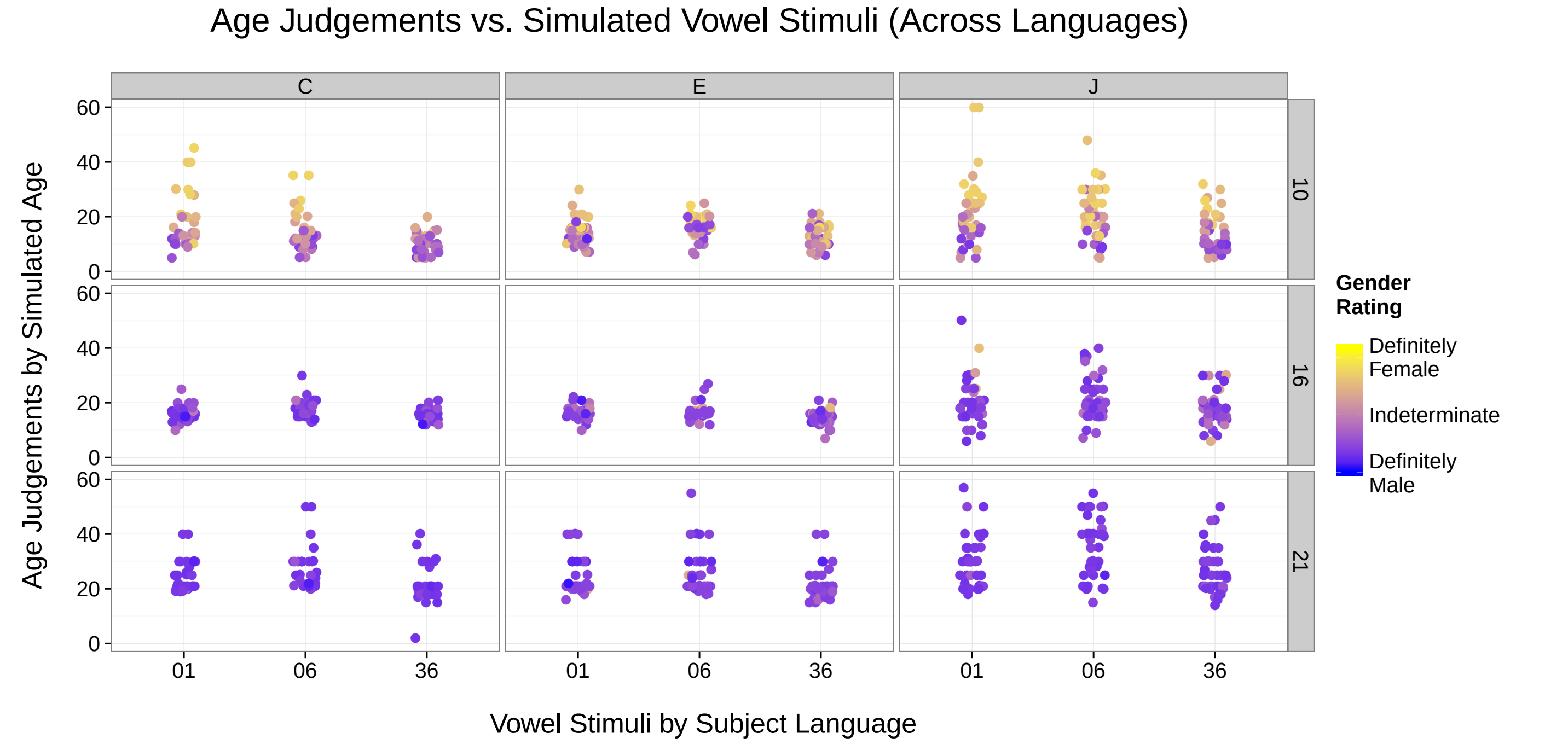
Pooled Age and Gender Responses





- Boxplots (median, interquartile range, and full range of values) for the age judgments (left) and the gender judgments (right) from all 57 listeners to all 6 trials for each vocal tract age.
- The age judgements generally increase as the age of the VLAM increases, while the gender judgements demonstrate ambiguity in interpretation up to age 10, which is resolved by sexual dimorphism at the simulated age 16.

Intimations of Culture-specific Nonuniformity



- A set of column scatter plots arranged by vocal tract age and language, with data points grouped by vowel stimuli, for the age judgments from all 57 listeners to all 6 trials for vocal tract ages 10, 16, and 21 years, with the color indicating the assigned gender rating.
- The Japanese subjects are assigning an older rating to 21 year old vowels, and this is especially true for stimulus 6.

Discussion

- Age responses do increase as the simulated vocal tract age increases, but the interpretation of age is variable across listeners and interacts substantially with the interpretation of gender for the simulated 10-year-old.
- Gender judgments suggest a strong ambiguity up until age 10, at which point gender judgements become bimodal, and contingent on age judgement.
- Of course, this does not mean that natural language-specific stimuli would be as ambiguous, since there could well be language- (and culture-) specific cues to age and gender that are not available in these synthesized stimuli.
- For example, the much older responses to stimulus 6 in the Japanese speakers' judgements could be an artifact of the quality of this stimulus, which is [u] rather than the unrounded [ɯ] that is the prototypical value for the high back vowel of Japanese.
- The rounded quality of the [u] might make it be heard as "careful speech" by Japanese listeners (see Okada, 1999), which could suggest an older, more deliberate talker.

Future work

- Exploring these language effects further may reveal the depth of influence of culture on aspects of perception, especially vowel normalization.
- While further experiments will be necessary to understand the full extent of language effects, the current preliminary results strongly support models of vowel normalization across different talker types as a learned process rather than an innate hard-wired one.

Corresponding author: Andrew R. Plummer - Department of Linguistics - The Ohio State University - Columbus, Ohio, USA

Mail: plummer@ling.ohio-state.edu

WWW: http://www.learningtotalk.org