How variable is variable phonetic input used in High Variability Phonetic Training paradigms?

Sarah Schaech

LING 2020

December 4, 2025

Introduction/Background

- L2 speech perception is difficult: Perception influenced by L1 sound system (Flege & Bohn, 2021; Best & Tyler, 2007)
- Perceptual training effective at improving listeners' ability to discriminate non-native sounds (Sakai & Moorman, 2018)
- High variability phonetic training (HVPT): highly effective L2 perceptual training method (Thomson, 2018; Barriuso & Hayes-Harb, 2018; Uchihara et al., 2025)
  - High variability = input: Repetitions of target sound presented in multiple talker voices and multiple phonetic contexts (e.g. surrounding phones), FCID task
  - Idea is that variability helps learners "tune in" to acoustic info that defines a sound/cues a contrast and "tune out" irrelevant information (Barriuso & Hayes-Harb, 2018)

Research Question

- Should we claim that input used in HVPT is "highly variable" without performing an acoustic analysis?
- How different are segments produced by many different talkers and presented in many different contexts from each other?
- **To what extent are stimuli in an HVPT paradigm different from each other...**
  - by target sound VOT between talkers and contexts?
  - by neighboring vowel formant frequencies?
- Current analyses:
  - Compare each segment VOT for each talker by condition
  - Compare vowel space (vowel following target) for each segment by condition

HVPT Stimuli

- Acoustically analyzed 80 unique tokens of stimuli used in a recent HVPT study (Nagle et al., 2025)
  - They investigated impacts of 2- vs 6-talker training on Spanish /p/-/b/ for L1 English speakers

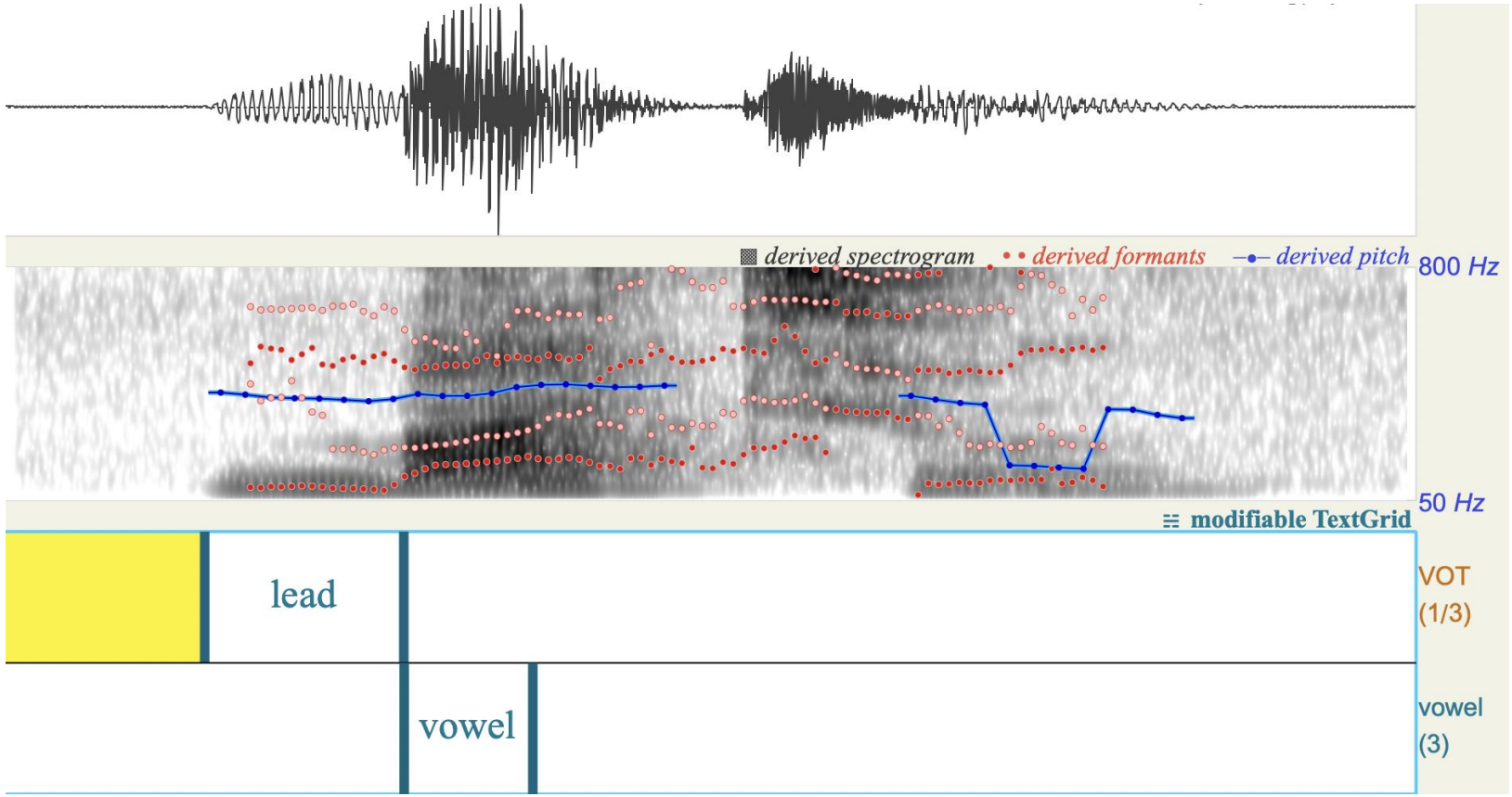
6-talker condition	2-talker condition
10 words x 6 talkers = 60 tokens	20 words x 2 talkers = 40 tokens
/p/ or /b/ + /a e i o u/ + /s n m/	/p/ or /b/ + /a e i o u/ + /s n m l tʃ/
Ex: "baso"	Ex: "picho"

Methods: Process

1. Measure acoustic values (VOT, F1, F2) in Praat
2. Extract values from Praat output, import in R
3. Combine all in a dataframe
4. Carry out analyses

Methods: Step 1

- Praat measurements and extraction



Methods: Step 2

- Praat output -> R
  - Read filepaths for labels of each token (sound file/TextGrid)
  - Use rPraat package to get VOT duration from each TextGrid -> df with filepaths
  - Separate filepath names by regex to extract word, voicing, sex, and talker ID variables
  - Praat script was necessary (at this point in time) to extract F1 and F2 info from the sound files
    - Read this output in and left joined with the VOT duration dataframe (Step 3: see next slide)
    - Created a key for duration and formant dataframes by using rownumber()

```
# extract variables from audio file names
all_data <- all_data |>
  separate_wider_regex(
    file,
    patterns = c(
      word = "^[bp].+",
      "_",
      voicing = "voiced|voiceless",
      "_",
      sex = ".",
      "_",
      talker_id = "\\d+",
      "[_.]",
      rest = "\\S+")
  )
```

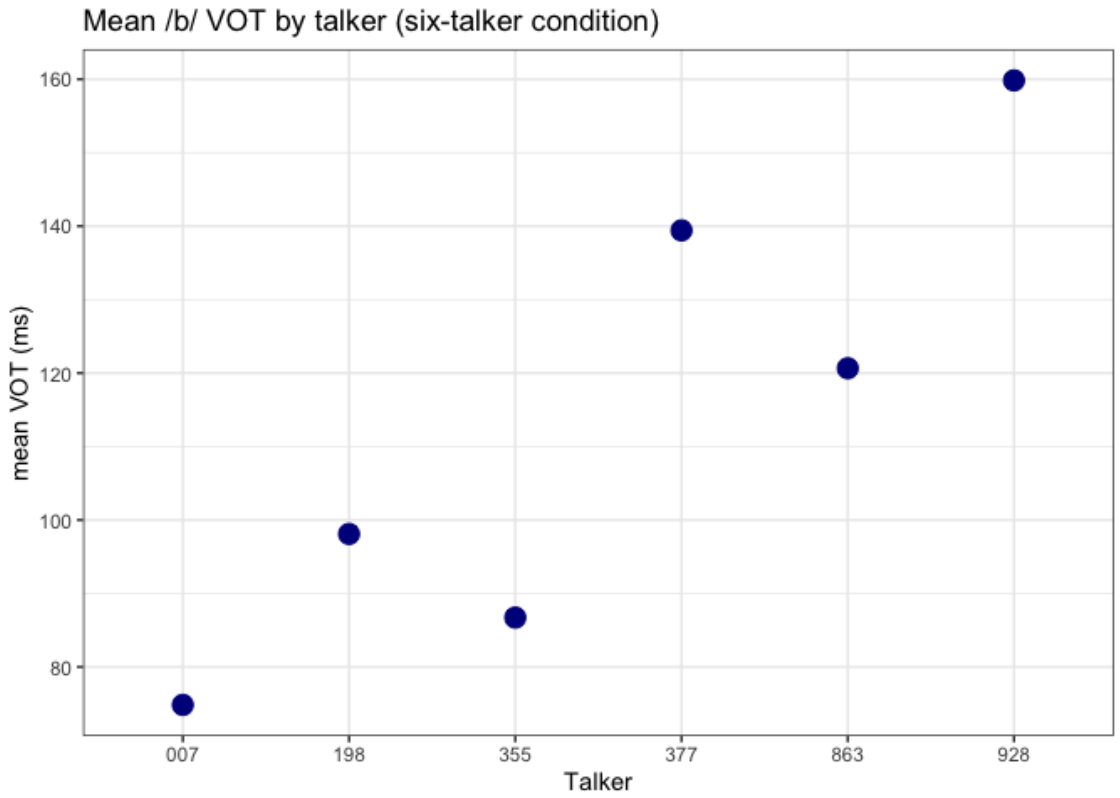
Methods: Step 3

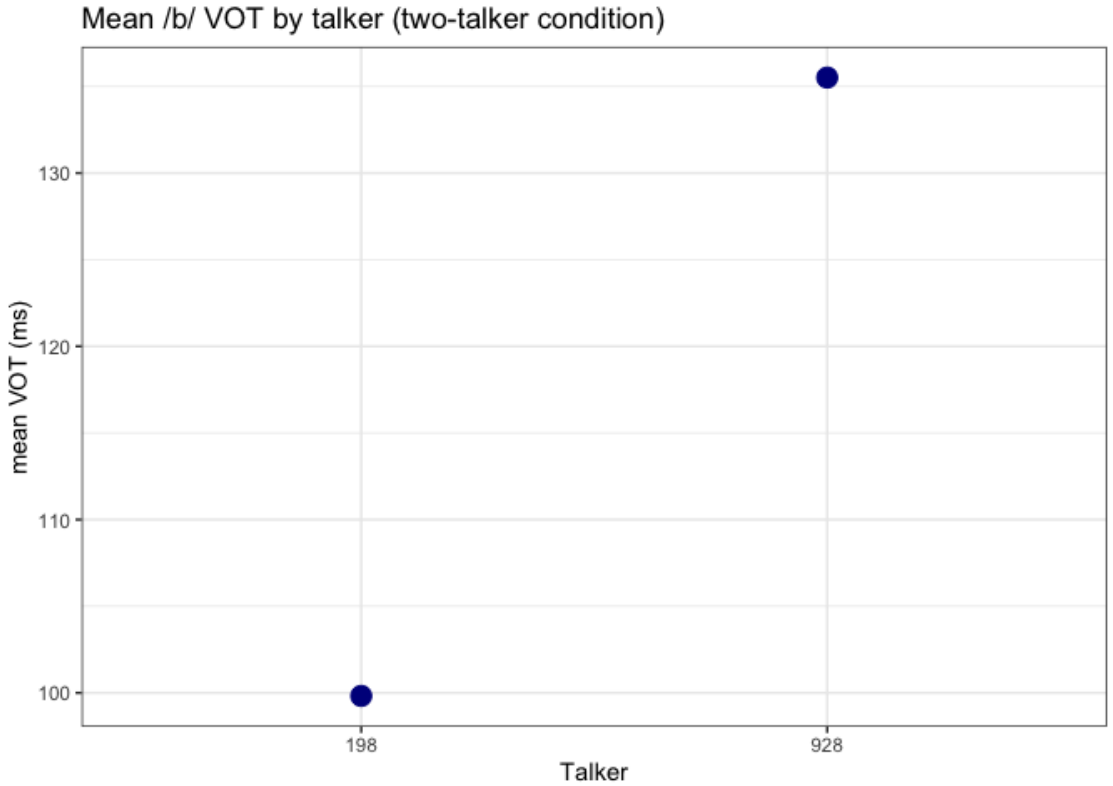
- Combine VOT duration data + vowel formants data

	word	voicing	sex	talker_id	f1_mid	f1_mean	f2_mid	f2_mean	f3_mid	f3_mean	vot_duration_ms	two_talker	six_talker
1	bacho	voiced	f	928	796	797	1288	1297	2916	2972	121.167226	y	n
2	bacho	voiced	m	198	625	590	1285	1298	2490	2480	85.526981	y	n
3	baso	voiced	f	007	773	761	1176	1209	2432	2421	22.103489	n	y
4	baso	voiced	f	863	1033	913	1396	1285	2782	2753	109.977731	n	y
5	baso	voiced	f	928	792	824	1243	1222	3076	3046	185.444701	y	y
6	baso	voiced	m	198	586	582	1192	1191	2434	2422	100.159883	y	y
7	baso	voiced	m	355	834	834	1288	1288	2961	2961	69.873922	n	y
8	baso	voiced	m	377	773	778	1059	1111	2616	2605	125.174844	n	y
9	belo	voiced	f	928	458	451	2156	2171	2848	2846	93.253077	y	n
10	belo	voiced	m	198	443	435	1665	1703	2345	2348	94.739298	y	n

Results/Analysis

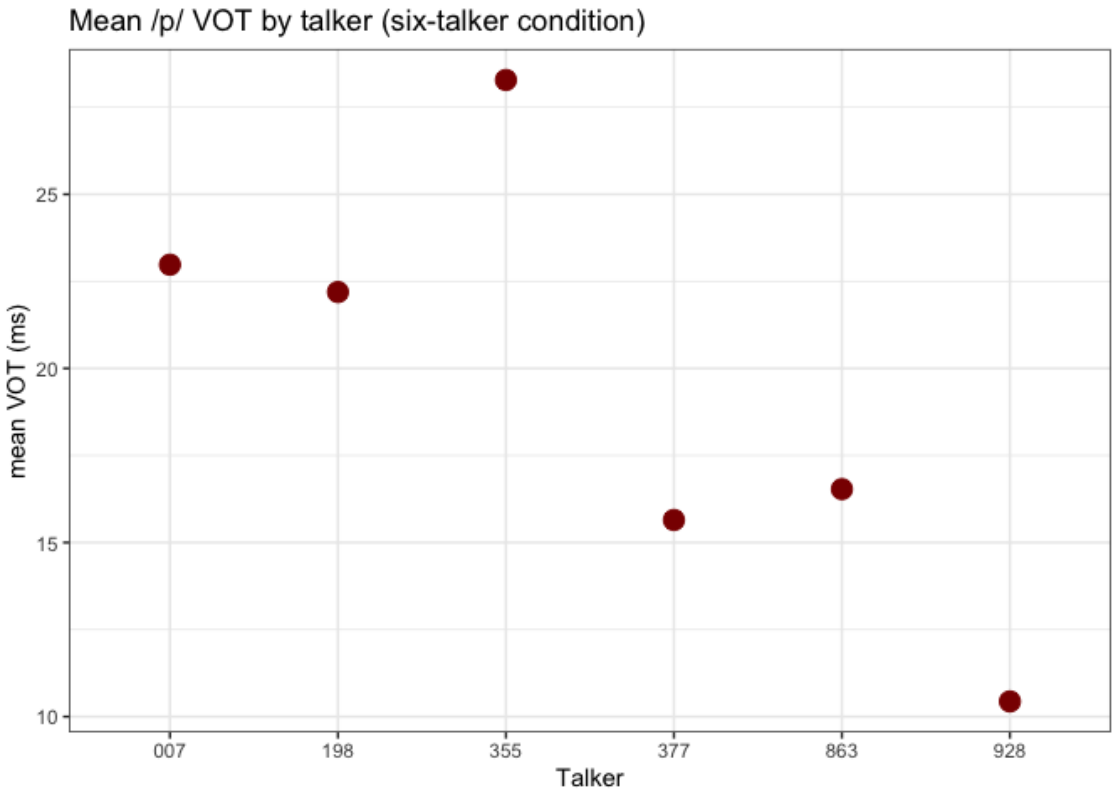
- Mean VOT for /b/ by talker for six-talker condition vs two-talker condition

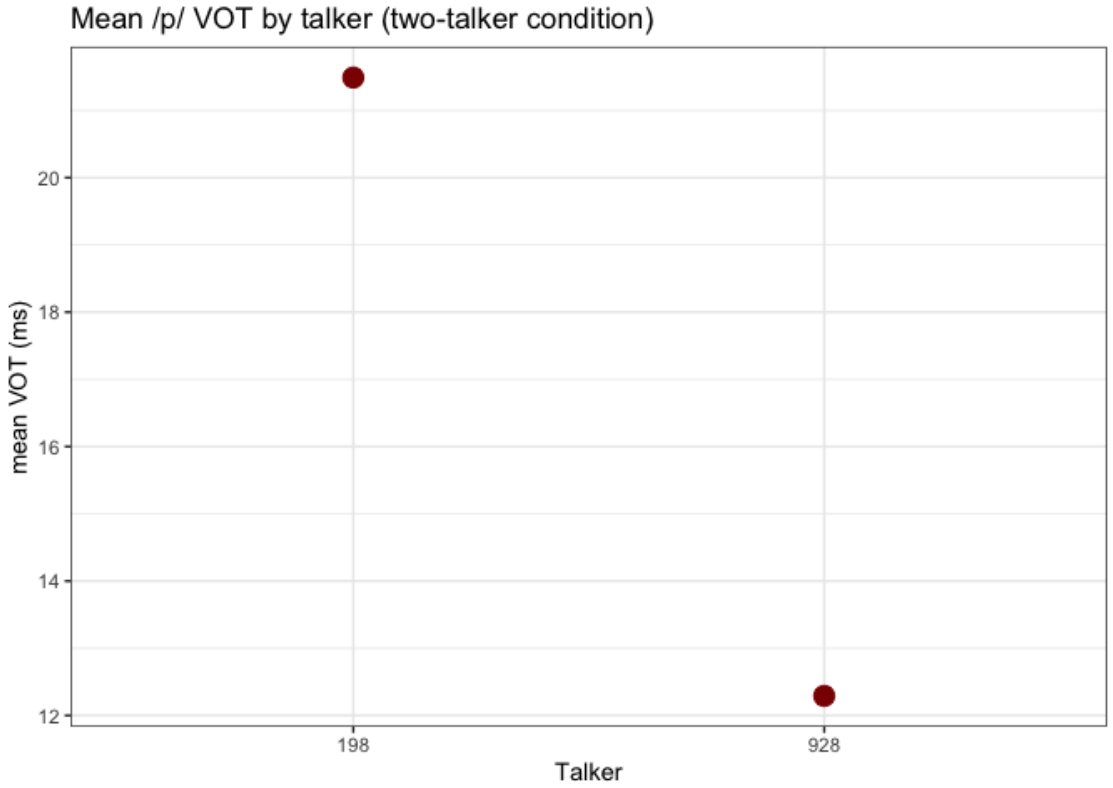




Results/Analysis

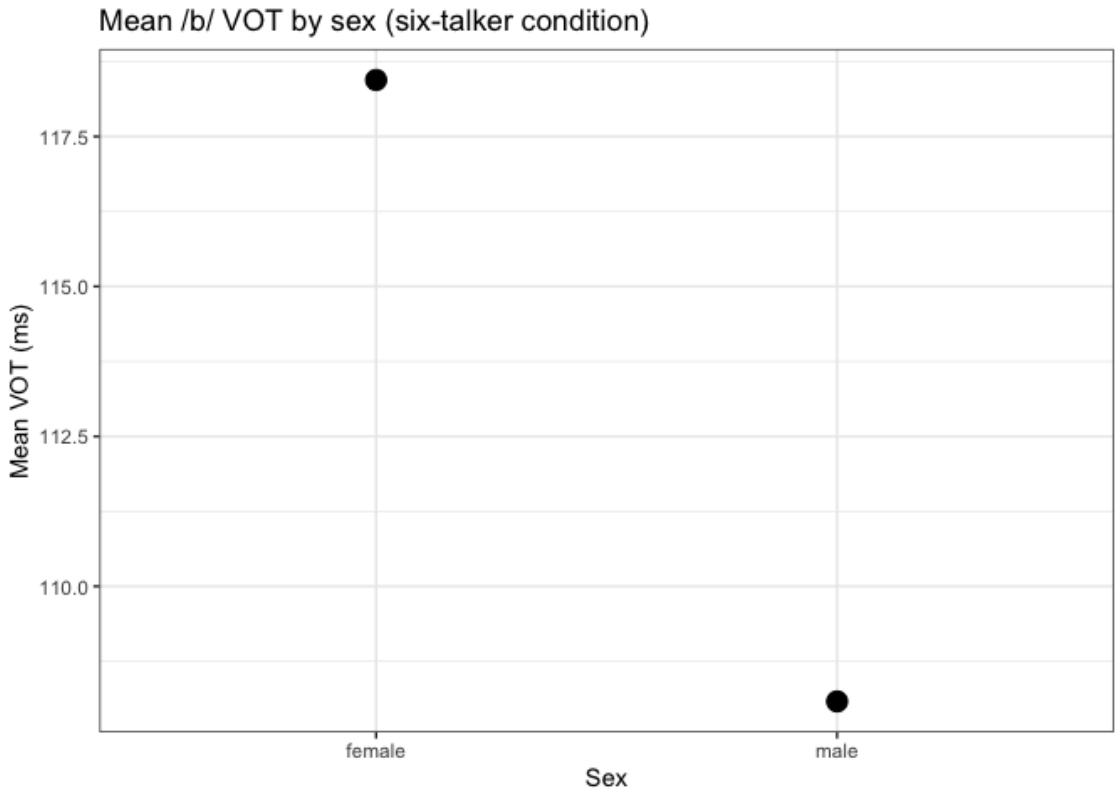
- Mean VOT for /p/ by talker for six-talker condition vs two-talker condition

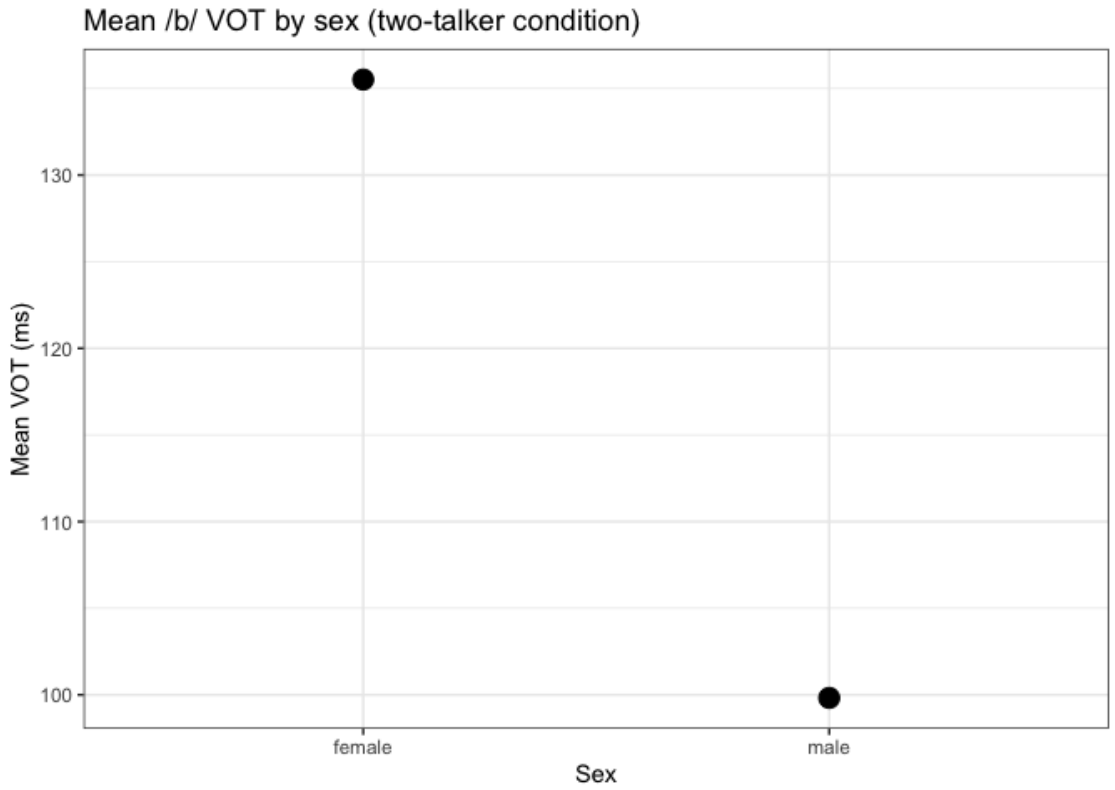




Results/Analysis

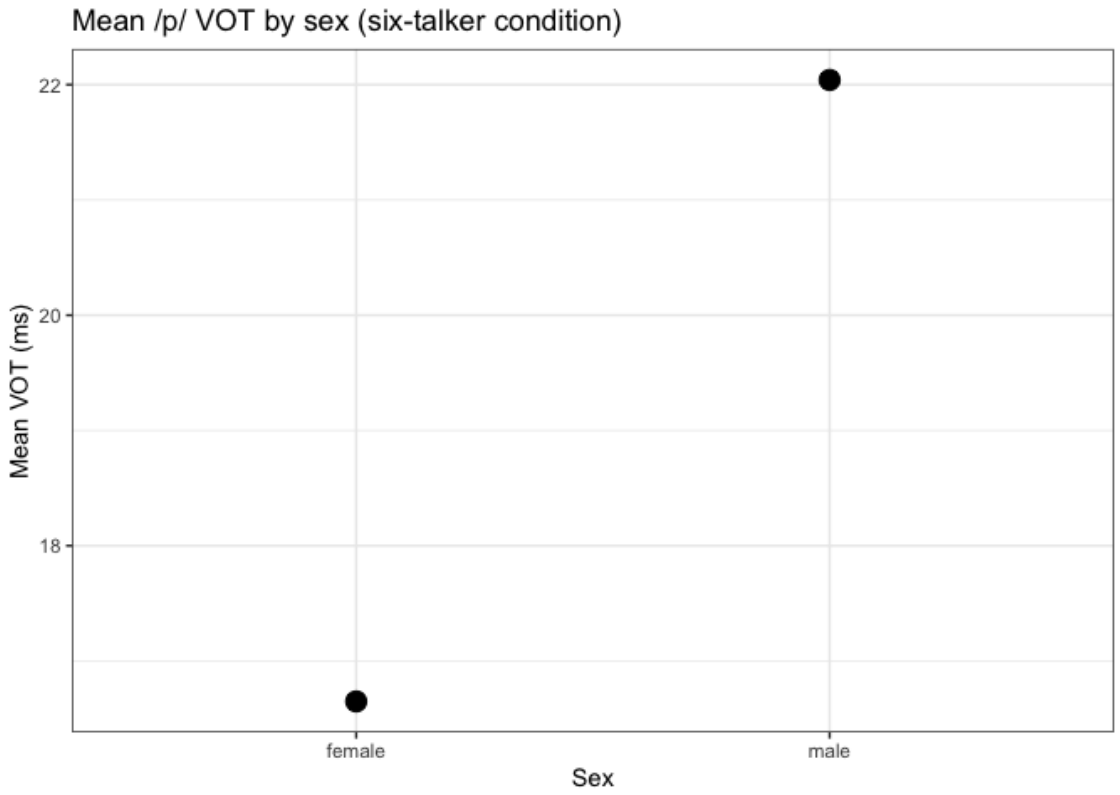
- Mean VOT for /b/ by sex for six-talker condition vs two-talker condition

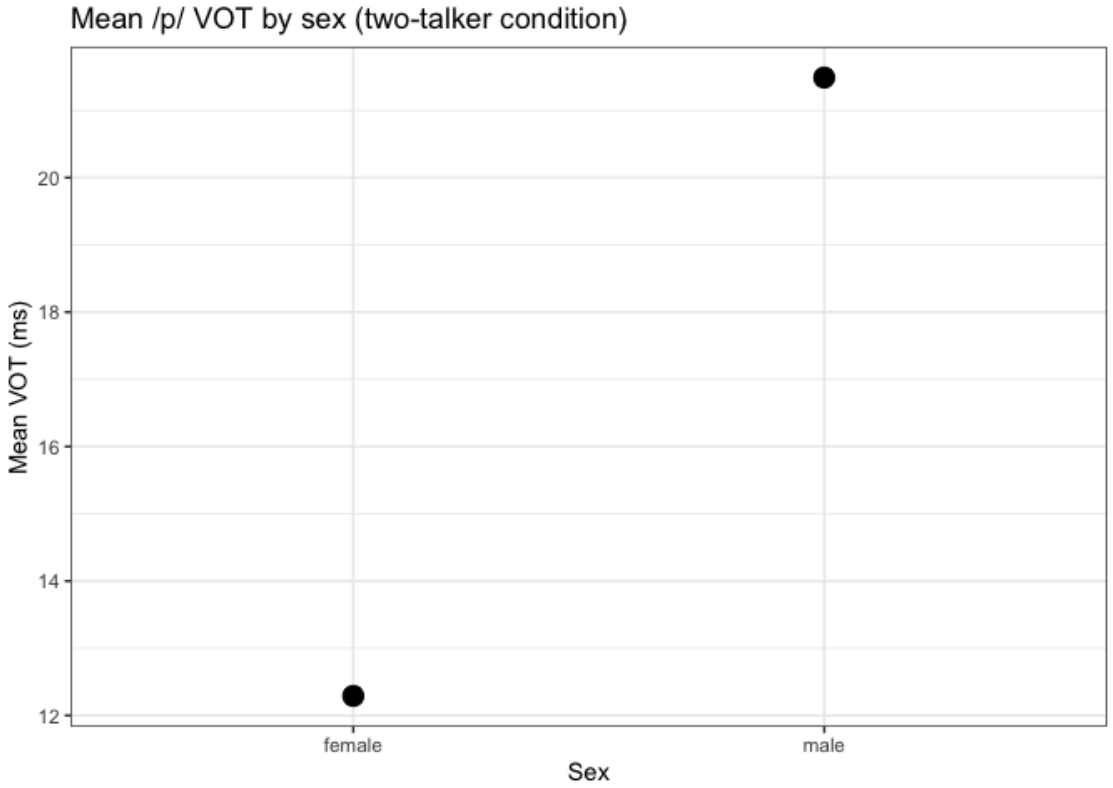




Results/Analysis

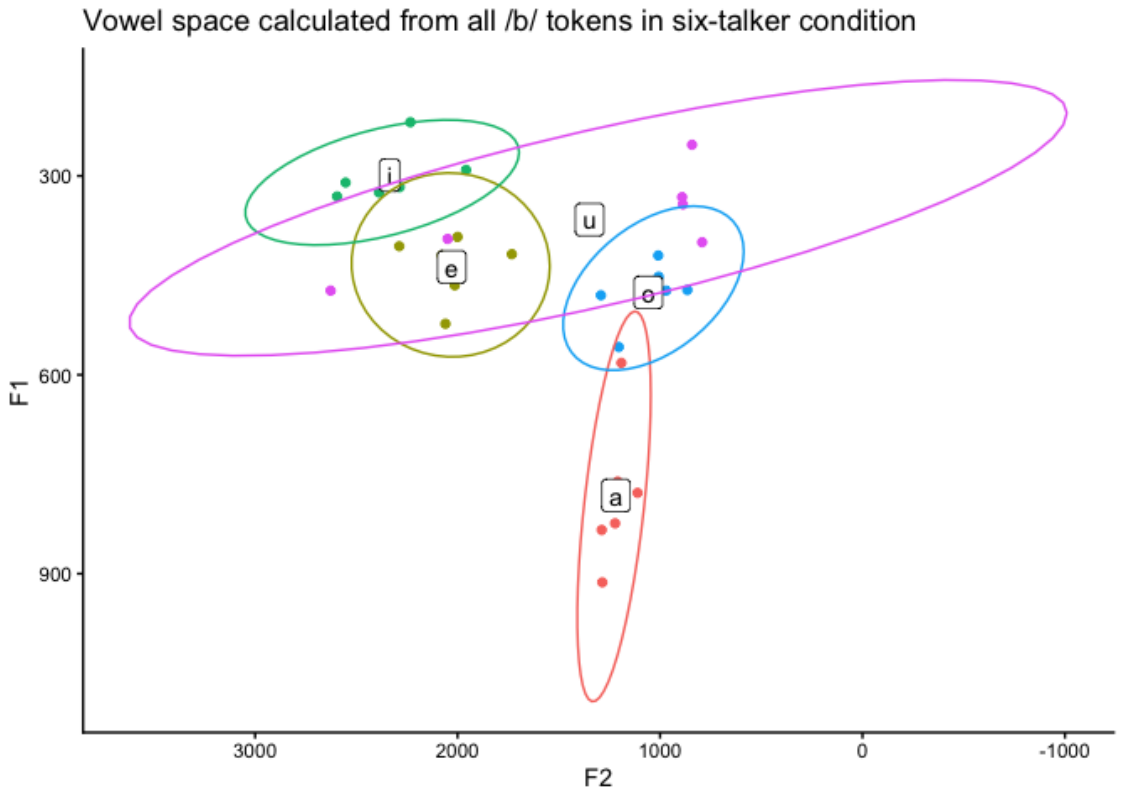
- Mean VOT for /p/ by sex for six-talker condition vs two-talker condition

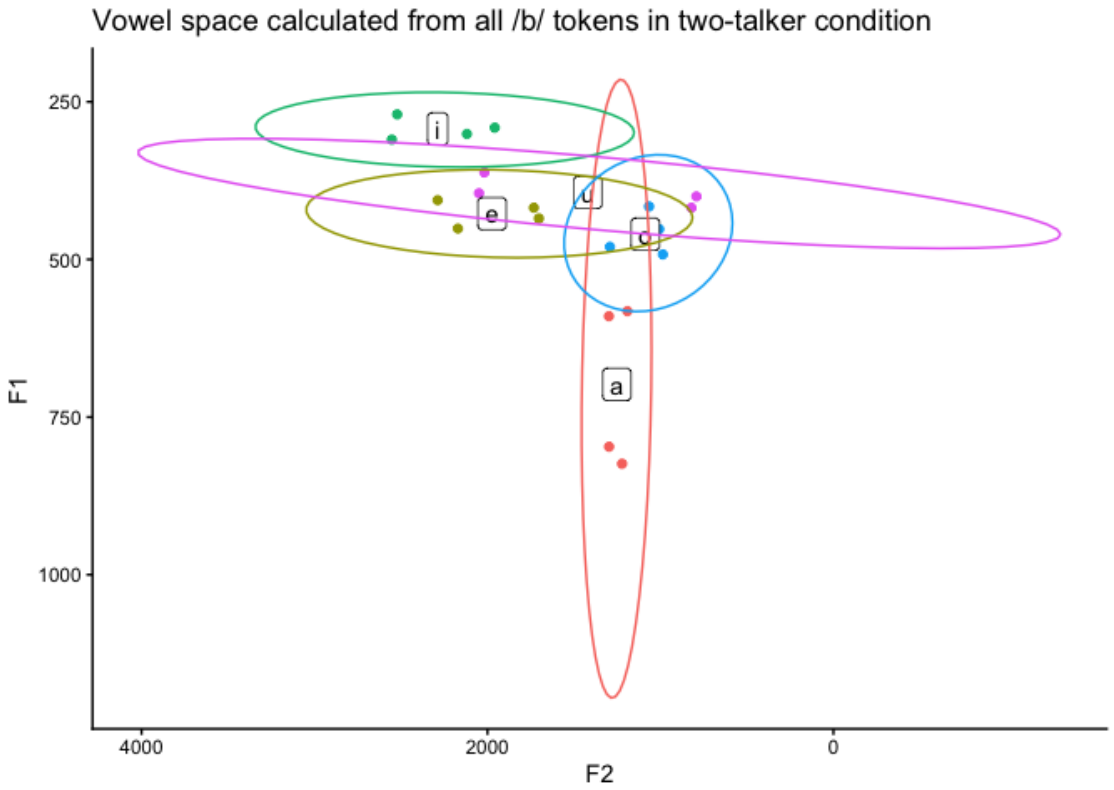




Results/Analysis

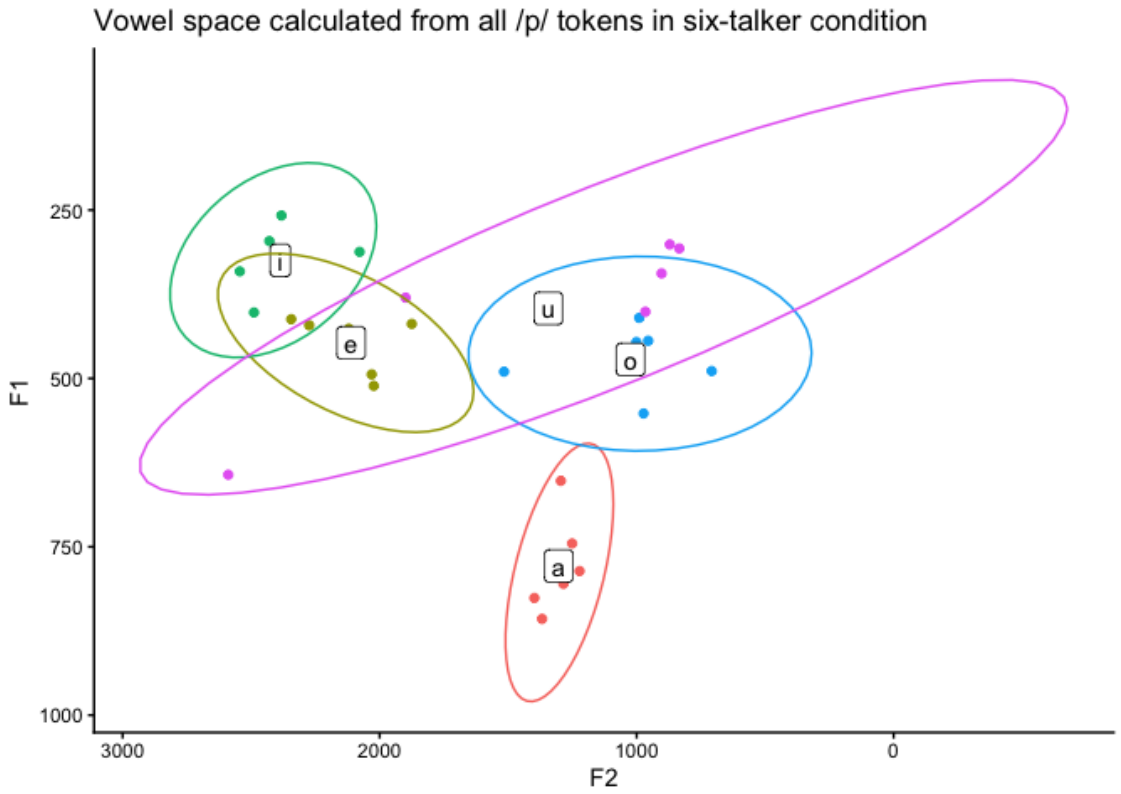
- Vowel space for /b/ across talkers for six-talker condition vs two-talker condition

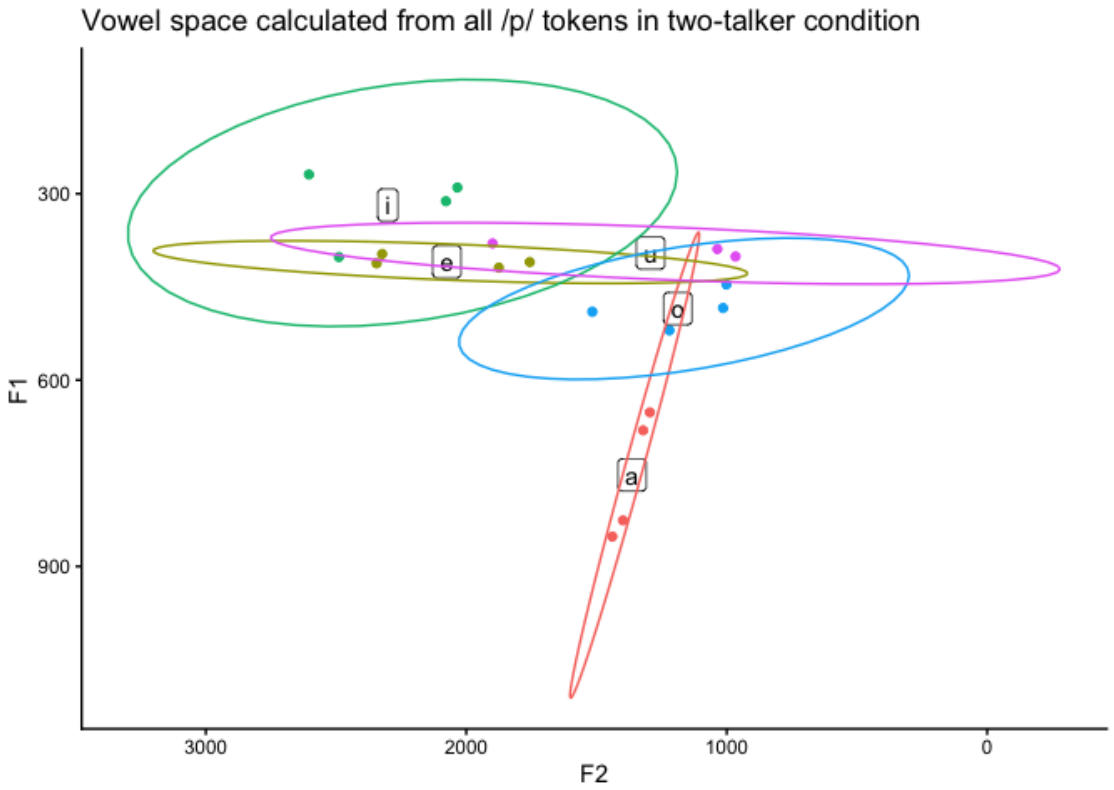




Results/Analysis

- Vowel space for /p/ across talkers for six-talker condition vs two-talker condition





Discussion

- **To what extent are stimuli in an HVPT paradigm different from each other…**
  - by target sound VOT between talkers and contexts?
  - by neighboring vowel formant frequencies?
- Preliminary interpretation:
  - If there was not much variability between talkers, we would see data points much closer together…but for each value (VOT and vowel formants), plenty of dispersion is observable
  - Interpretation is difficult:
    - What is the baseline?
    - Two-talker condition actually included 5 extra phonetic contexts (adds input variability?)
  - Application: The variability is there, but we're still figuring out how *much* is optimal
- Future work (for this project):
  - Tweak plots to make differences in dispersion more visually detectable
  - Consider other variables

References

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, *30*(1).

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). https://doi.org/10.1075/lllt.17.07bes

Flege, J., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). https://doi.org/10.1017/9781108886901.002

Nagle, C., Bruun, S., & Zárate-Sández, G. (2025). Comparing lower and higher variability multi-talker perceptual training. *Applied Psycholinguistics*, *46*, e14, Article e14. https://doi.org/10.1017/S0142716425000141

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224. https://doi.org/10.1017/S0142716417000418

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT). *Journal of Second Language Pronunciation*, *4*(2), 208-231. https://doi.org/https://doi.org/10.1075/jslp.17038.tho

Uchihara, T., Karas, M., & Thomson, R. I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*, *47*(3), 794-827. https://doi.org/10.1017/S0272263125100879