Effects of talker variability on categorization of spectrally degraded vowels

Emily A. Dickey and Christian E. Stilp
Department of Psychological and Brain Sciences, University of Louisville

INTRODUCTION

- Perception of a speech sound is heavily influenced by surrounding sounds. When spectral properties differ between earlier (context) and later (target) sounds, this can produce **spectral contrast effects (SCEs)** that bias categorization of later sounds.

Context	More likely to perceive
Sentence (unmodified)	"ih" or "eh"
Sentence with "eh"-like (high F_1) frequencies emphasized	"ih" (low F_1)
Sentence with "ih"-like (low F_1) frequencies emphasized	"eh" (high F_1)

- SCEs affect vowel categorization in simulations of cochlear implant (CI) processing (Stilp, 2017) and for CI users (Feng & Oxenham, 2018).
- Also, normal-hearing listeners' SCEs are smaller when they hear 200 different talkers compared to one talker (Assgari & Stilp, 2015).
- Finally, CI users can struggle with talker discrimination (e.g., Fu et al., 2004; Massida et al., 2011; Stickney et al., 2004).
- This study tested the intersection of these three threads. We predicted that talker variability would not affect SCE magnitudes in simulated CI (noise-vocoded) vowel categorization.

METHODS

Participants

40 native English undergraduate students with self-reported normal hearing

Stimuli

1. Context Sentences

- 1 male talker reading 200 different sentences (HINT: Nilsson et al., 1994); 200 talkers reading 200 different sentences (TIMIT: Garofolo et al., 1990)
- Filtered to amplify low-F_1 frequencies (100-400 Hz) or high-F_1 frequencies (550-850 Hz) by 5 dB to produce SCEs (after Assgari & Stilp, 2015)
- Noise-vocoded from 100-5000 Hz using 4, 8, 12, or 24 channels

2. Target Vowels

- 10-step series perceptually varying from "ih" (as in 'bit') to "eh" (as in 'bet'); as tested in Assgari & Stilp (2015)

Procedure

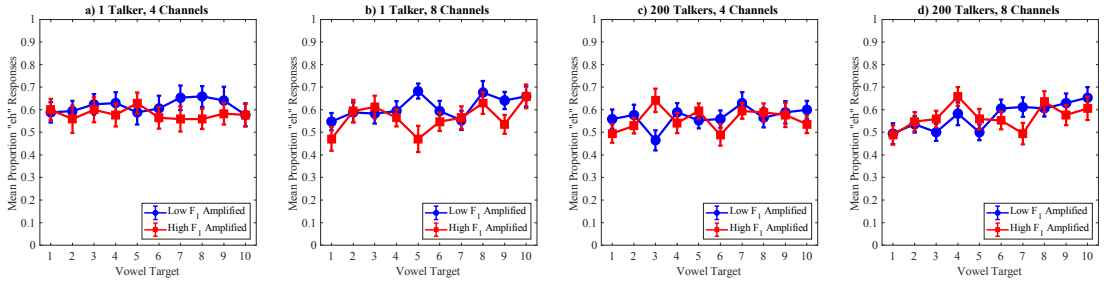
- On each trial, participants heard a sentence and categorized the following target vowel
- 4 blocks of 200 trials tested in counterbalanced orders
  - Expt. 1: spectral channels (4 / 8) x number of talkers (1 / 200)
  - Expt. 2: spectral channels (12 / 24) x number of talkers (1 / 200)
- Following Stilp (2017), a participant's data were removed if s/he relied heavily on one response category (>80% of responses) in the easiest condition tested (1 talker / higher number of channels tested)
  - Expt. 1: n = 3 removed; Expt. 2: n = 2 removed

RESULTS

SCE Calculation

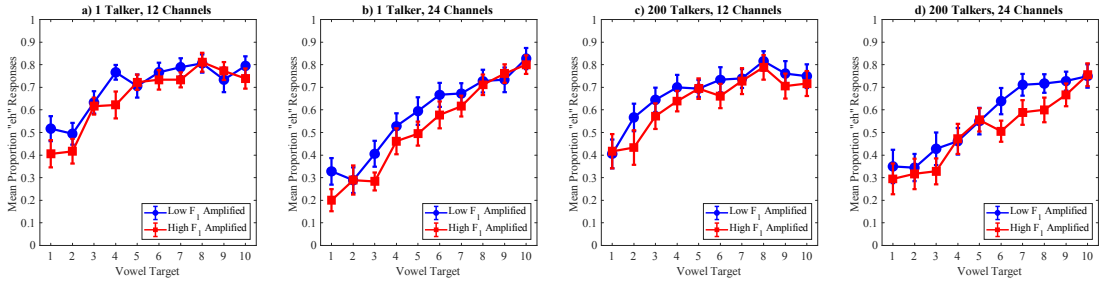
- In each block, SCEs were measured as the difference in mean percentage "eh" responses across filtering conditions (low-F_1-amplified versus high-F_1-amplified)

Experiment 1 (n = 17)

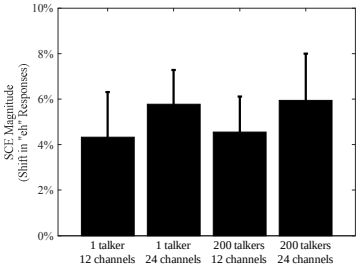


- Experiment 1 was more challenging than anticipated. Response curves were flat lines (means and standard errors shown above), reflecting participants' inability to distinguish target vowels. Measuring categorization shifts (SCEs) due to preceding spectral context is problematic when listeners do not have clear categories to begin with. This led to Experiment 2.

Experiment 2 (n = 18)



- Performance improved substantially in Experiment 2 owing to testing higher spectral resolution (more channels than Experiment 1; response means and standard errors shown above)
- SCEs were calculated for each participant in each block (means and standard errors shown at right), then analyzed in a 2 (number of talkers) by 2 (number of spectral channels) repeated measures ANOVA
  - No main effect of number of talkers: $F(1,17) = 0.03$, $p = 0.88$
  - No main effect of number of channels: $F(1,17) = 0.57$, $p = 0.46$
  - No interaction between talkers and channels: $F(1,17) = 0.00$, $p = 0.97$



DISCUSSION

- The interaction between the number of talkers and the number of spectral channels in noise vocoding was not statistically significant. SCE magnitudes were relatively constant across conditions. This supports the prediction that talker variability would not affect SCE magnitudes. However, the challenge of interpreting a null result remains.
- Results may be due in part to impoverished pitch cues in noise vocoding. Assgari et al. (2019) showed that variability of acoustically dissimilar talkers (high f0 variability) resulted in smaller SCEs, but variability of acoustically similar talkers (low f0 variability) did not. In noise vocoding and CI processing, impoverished pitch information would contribute to talkers sounding less distinct from one another, producing results like those observed here.
- Stilp (2017) reported clear SCEs for stimuli vocoded with 6 spectral channels, but that was not the case for 4 and 8 channels here. The big difference between studies is that Stilp (2017) tested the same context sentence by the same talker on every trial. Here, there was always a different sentence on each trial. Talker and / or stimulus variability might have made this experiment more challenging, producing flat response curves.

Open Questions

- How many vocoder spectral channels should be tested in experiments like these? A wide range was tested here to reflect individual differences among CI users, but this modulated task difficulty and the interpretability of SCEs.
- Should there have been a practice session before testing blocks? This was not a training / perceptual learning study, but perhaps practice with feedback would have produced clearer results in Experiment 1 (and perhaps Experiment 2 as well).

Conclusion

- Talker variability does not appear to affect spectral context effects in categorization of noise-vocoded vowels. This patterns differently from talker variability affecting SCEs for normal-hearing listeners.

REFERENCES

Assgari AA, Stilp CE (2015) *JASA, 138(5)*, 3023-3032
Assgari AA, Theodore RM, Stilp CE (2019) *JASA, 145(3)*, 1443-1454
Feng L, Oxenham AJ (2018) *JASA, 143(6)*, EL468-EL473
Fu Q-J, Chinchilla S, Galvin III JJ (2004) *JARO, 5(3)*, 253-260
Garofolo JS et al. (1990) DARPA TIMIT CD-ROM. *NIST.*
Massida Z, Belin P, James C, Rouger J, Fraysse B, Barone P, Deguine O (2011) *Hear Res, 275(1-2)*, 120-129
Nilsson M, Soli SD, Sullivan JA (1994) *JASA 95(2)*, 1085-1099
Stickney GS, Zeng F-G, Litovsky RY, Assmann PF (2004) *JASA 116(2)*, 1081-1091
Stilp CE (2017) *JARO, 18(3)*, 465-481