Trial-to-trial variability in talkers' fundamental frequencies restrains spectral context effects in vowel categorization

Ashley A. Assgari and Christian E. Stilp

Department of Psychological and Brain Sciences, University of Louisville

INTRODUCTION

Spectral contrast effects (SCEs) occur when the auditory system perceptually magnifies spectral differences between sounds:

Precursor	More likely to hear
Sentence (unmodified)	/ɪ/ or /ɛ/ vowel target
Sentence with /ɛ/-like (high F_1) frequencies emphasized	/ɪ/ (low F_1)
Sentence with /ɪ/-like (low F_1) frequencies emphasized	/ɛ/ (high F_1)

But, these context effects were significantly smaller when precursor sentences were spoken by 200 different talkers versus a single talker (Assgari & Stilp, 2015). This was due to variability in mean f0: smaller SCEs when variability in precursor sentences' mean f0 was high, normal-sized SCEs when mean f0 variability was low (Assgari et al., under review)

Many talker normalization studies found that speech perception is slower and/or less accurate when hearing multiple talkers
- In particular, perception is less accurate when talkers are acoustically different than when they are acoustically similar (Goldinger, 1996)

Randomizing talkers introduces high trial-to-trial (local) variability and high session-level (global) variability. Was local or global variability responsible for smaller SCEs?

Here, we investigated whether local or global variability in talkers' mean f0s changes SCEs and response times

STIMULI

Sentences
- The same 40 high-f0-variability sentences previously used in Assgari et al. (under review)
- Measured sentences' mean f0s in Praat
- Bandpass filter added +5 dB spectral peaks to sentences in low-F_1 (100-400 Hz) or high-F_1 (550-850 Hz) region

Vowels
- The same morphed natural vowels from [ɪ] to [ɛ] previously used in Assgari et al. (under review)

## PROCEDURE AND ANALYSES

Procedure
- Each trial presented a sentence then a target vowel, which listeners identified as "'ih' as in bit" or "'eh' as in bet" via a response box
- *Practice:* 20 sentences from the AzBio corpus (Spahr et al., 2012) paired with endpoint vowels; >80% accuracy needed to continue
- *Test:* 4 blocks of 160 trials (each talker repeated 4 times per block in multi-talker conditions)
  - Single: One sentence from one talker previously used in Assgari & Stilp (2015)
  - Ascending: Trials arranged from lowest to highest mean f0
  - Descending: Trials arranged from highest to lowest mean f0
  - Maximum Variability: Trials arranged to maximize successive differences in mean f0 (lowest-f0 man first, then lowest-f0 woman, then next-lowest-f0 man, etc.)


High f0 Variability Sentences

SCE Calculation
- Logistic regressions were fit to each listener's responses following low-F_1 and high-F_1 sentences
- 50% points were calculated from each regression equation
- SCE = the difference in 50% points between low-F_1 and high-F_1 functions (i.e., number of stimulus steps along the abscissa; see right)

Response Times
- Measured starting at vowel onset
- Responses <150 ms and >3 SDs from each individual's average removed
- Averaged across low-F_1 and high-F_1 filtering conditions

SCE = 0.8 stimulus steps

RESULTS (n=21)
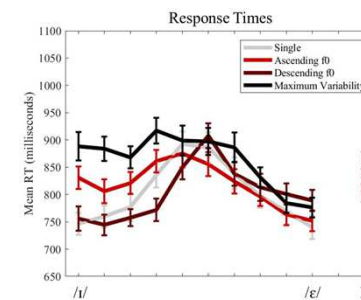

SCE Magnitudes
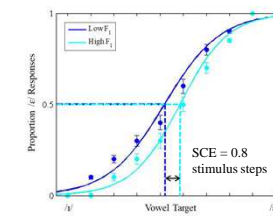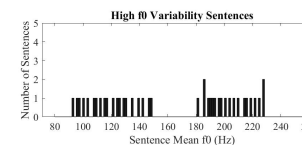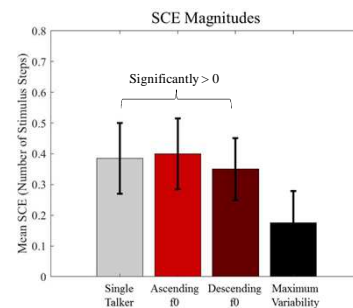
Significantly > 0

Repeated Measures ANOVA:
- No significant differences; $F_{3,60} = 0.91$, $p = 0.44$
- But, type II error is a possibility when 3 groups have equivalent means

Did an SCE occur? (one-way t-tests against 0)
- SCEs in Single, Ascending, Descending (p's < 0.01)
- No SCE in Maximum Variability ($p = 0.11$)

Response Times

Repeated Measures ANOVA:
Condition: $F_{3,60} = 3.74$, $p = 0.02$, $\eta_p^2 = 0.16$
- Maximum > Single & Descending (p's < 0.05)
Vowel: $F_{3.04,60.72} = 13.77$, $p < 0.01$, $\eta_p^2 = 0.41$
- Slowest for mid-continuum
- Slower at /ɪ/ end of continuum than /ɛ/ end
Interaction: $F_{27,540} = 4.24$, $p < 0.01$, $\eta_p^2 = 0.18$
- Biggest condition differences at /ɪ/ end

DISCUSSION

When talkers on successive trials are acoustically variable, speech perception is slower and spectral context biases categorization less

When talkers on successive trials are acoustically similar, response times and context effects are similar to single-talker conditions
- No differences in SCE magnitudes or response times for Ascending/Descending f0 conditions compared to the Single talker condition

Local (trial-to-trial) variability influences speech perception more than global (session-level) variability, which was equivalent across Ascending, Descending, and Maximum Variability conditions

Limitations and Future Directions
1) Predictability
- Trials in the Maximum Variability condition formed a pattern (low f0, high f0, low f0, high f0, ...). Could this undermine the effects of acoustic variability?
- Randomizing talkers may produce similar trial-to-trial variability without being predictable

2) Talker familiarity
- Talkers in multi-talker conditions were repeated 4x in each block; were listeners becoming familiar with them?
- More talkers with fewer repetitions should eliminate this possibility (e.g., 200 different talkers in Assgari & Stilp, 2015)

Conclusions
Perceptual costs incurred by hearing different talkers may be due to local (trial-to-trial) variability, not global variability

REFERENCES

Assgari, A. A. (2018). Assessing the Relationship between Talker Normalization and Spectral Contrast Effects in Speech Perception. (Dissertation), University of Louisville.

Assgari, A.A., & Stilp, C.E. (2015). Talker information influences spectral contrast effects in speech categorization. *Journal of the Acoustical Society of America*, 138(5), 3023-3032.

Assgari, A.A., Theodore, R.M., & Stilp, C.E. (under review). Differential effects of talker variability on context effects in speech perception. *Journal of the Acoustical Society of America*.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., ... & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing*, 33(1), 112.