Effects of type, token, and talker variability in speech processing efficiency

Alexandra M. Kapadia, Jessica A.A. Tin, & Tyler K. Perrachione
Department of Speech, Language & Hearing Sciences, Boston University
PDF of this poster can be found at http://sites.bu.edu/cnrlab/publications/conference-presentations-abstracts/

BOSTON UNIVERSITY

Summary

- Phonetic variability across talkers imposes additional processing costs; however, within-talker phonetic variation is another, relatively unexplored source of variability in speech.
- It is unknown how processing costs from within-talker variation compare to those from between-talker variation, and how these different effects scale and interact [5, 10].
- Conditions factorially manipulated three dimensions of variability: **number of word choices (type)**, **number of talkers (talker)**, and **number of talker-specific exemplars per word (token)**.
- Participants performed a speeded word identification task with reaction time (RT) as the dependent variable [1, 8].
- Across all eight experimental levels, larger decision spaces (more target word choices) led to slower word identification.
- Word identification was also slower in conditions with mixed talkers and conditions with multiple exemplars.
- However, performance decrements due to talker variability were only present when variability in the other two dimensions was low, but decrements due to exemplar variability were present under all conditions.

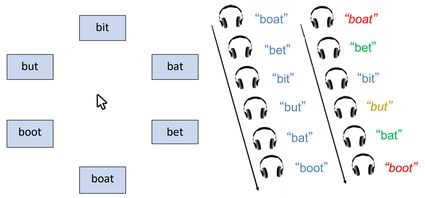
Methods

- **Stimuli:** 4 talkers (2 female, 2 male) each recorded 6 minimally-contrastive English words: bit, bet, bat, but, boat, boot → /ɪ/, /ɛ/, /æ/, /ʌ/, /o/, /u/.
  Token variability was elicited through 8 variations: (3 pitches x 2 durations + 2 contours).
- **Participants:** Native English speakers (*N* = 24; 18 female, 6 males; age 18-24 years).
- **Experiment:** Participants responded using a mouse with options presented on the screen. For two-word choice conditions, only the two relevant words were displayed.
- **Conditions:** Each combination of dimension values (low vs. high).

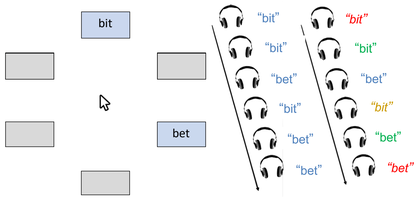
Condition	Talkers	Types	Tokens	Type Combinations	Trials
1	single	low	one	15	240
2	single	high	one	1	240
3	single	low	many	15	240
4	single	high	many	1	240
5	multi	low	one	15	240
6	multi	high	one	1	240
7	multi	low	many	15	240
8	multi	high	many	1	240

Degree of Variability: *Low* → *High*

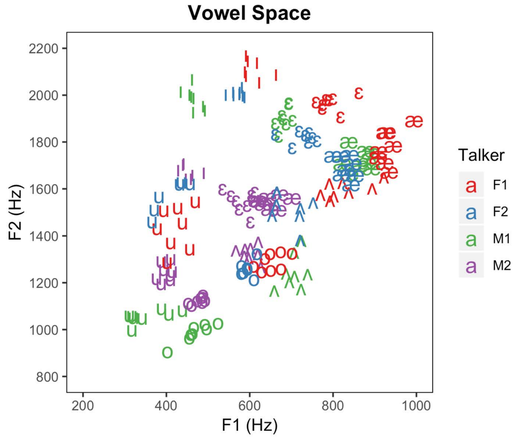
A. (L) Large decision space: six words. *(R)* Single-, mixed-talker conditions.



B. (L) Small decision space: two words. *(R)* Single-, mixed-talker conditions.



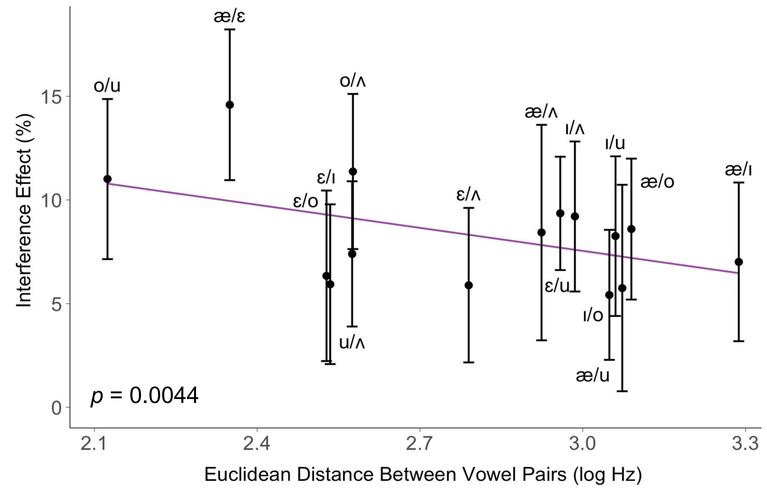
Stimulus Variability



- Phonetic variability plotted in F1 x F2 space across talkers for the words “bat”, “bet”, “bit”, “boat”, “boot”, and “but”.
- Many areas of acoustic-phonemic ambiguity, where vowel tokens for different categories overlap or are circumscribed.

Degree of Ambiguity

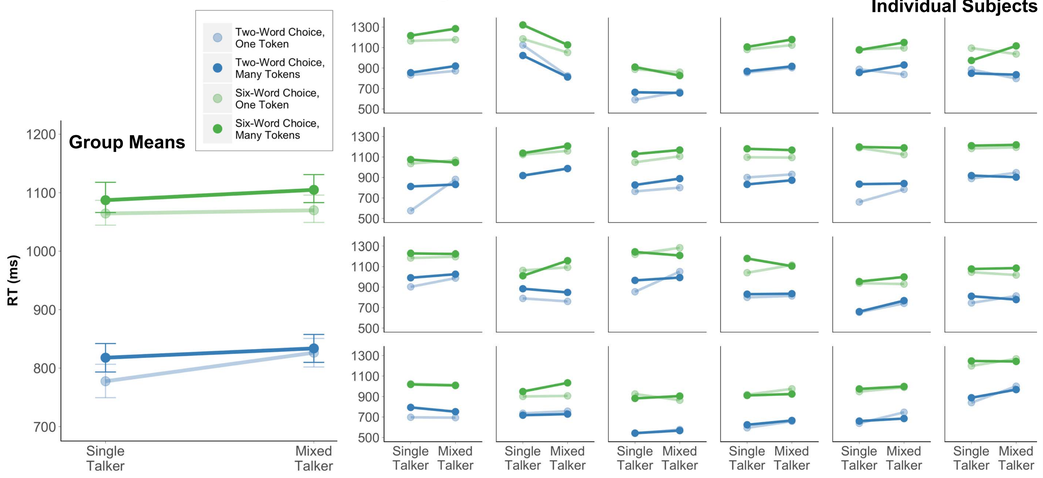
- Interference effect = [(mixed talker RT – single talker RT) / single talker RT * 100].
- Euclidean distance was calculated from the mean position of each vowel in F1 x F2 space using measurements from [4].
- Interference effect inversely scaled with the Euclidean distance (log Hz) between vowel pairs in the two-word choice (low-type) conditions.



Effects of Variability

- Significant three-way interaction between all factors: type (two- vs. six-word choice), talker (single vs. mixed), and token (one vs. many).
- Larger decision space (high type), exemplar variability (many tokens) led to slower reaction times (RT); talker variability (mixed talkers) also led to slower RTs only when other variability was low.

Effect/Interaction	*F*	*df(n, d)*	*p*
Talker (single vs. mixed)	6.53	(1, 23.6)	0.018
Type (low vs. high)	22.3	(1, 24.0)	2.2×10^{-16}
Token (one vs. many)	728.3	(1, 37.1)	3.3×10^{-5}
Talker x Type	13393.7	(1, 55.9)	8.1×10^{-14}
Talker x Token	12011.0	(1, 30.7)	3.1×10^{-8}
Type x Token	12285.2	(1, 24.4)	8.1×10^{-7}
Talker x Type x Token	59.3	(1, 12381.9)	1.5×10^{-14}



References

[1] Choi, J. Y., Hu, E. R., & Perrachione, T. K. 2018. Atten. Percept. Psychophys. 80, 784-797.
[2] Goldinger, S. D. 1996. J. Exp. Psychol. Learn. 22, 1166-1183.
[3] Green, K. P., Tomiak, G. R., & Kuhl, P. K. 1997. Percept. Psychophys. 59, 675-692.
[4] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. 1995. J. Acoust. Soc. Am. 97, 7099-3111.
[5] Kleinschmidt, D. F., & Jaeger, T. F. 2015. Psychol. Rev. 122, 148-203.
[6] Magnuson, J.S., & Nusbaum, H.C. 2007. J. Exp. Psychol. Human. 33, 391-409.
[7] McLennan, C. T., & Luce, P. A. (2005). J. Exp. Psychol. Learn. 31, 306.
[8] Mullennix, J.W., & Pisoni, D.B. 1990. Percept. Psychophys. 47, 379-390.
[9] Mullennix, J. W., Pisoni, D. B., & Martin, C. S. 1989. J. Acoust. Soc. Am. 85, 365-378.
[10] Newman, R. S., Clouse, S. A., & Burnham, J. L. 2001. J. Acoust. Soc. Am. 109, 1181-1196.
[11] Nygaard, L. C., Sommers, & M. S., Pisoni, D. B. (1995). Percept. Psychophys. 57, 989-1001.
[12] Sommers, M. S., & Barcroft, J. 2006. J. Acoust. Soc. Am. 119, 2406-2416.

Support

This research was supported by the NIDCD of the National Institutes of Health under award R03DC014045, and a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation to T.P. We thank Sung-Joo Lim and Melanie Matthies for their assistance.

BU Communication Neuroscience Research Laboratory
akapadia@bu.edu; tkp@bu.edu
http://sites.bu.edu/cnrlab/