

Introduction

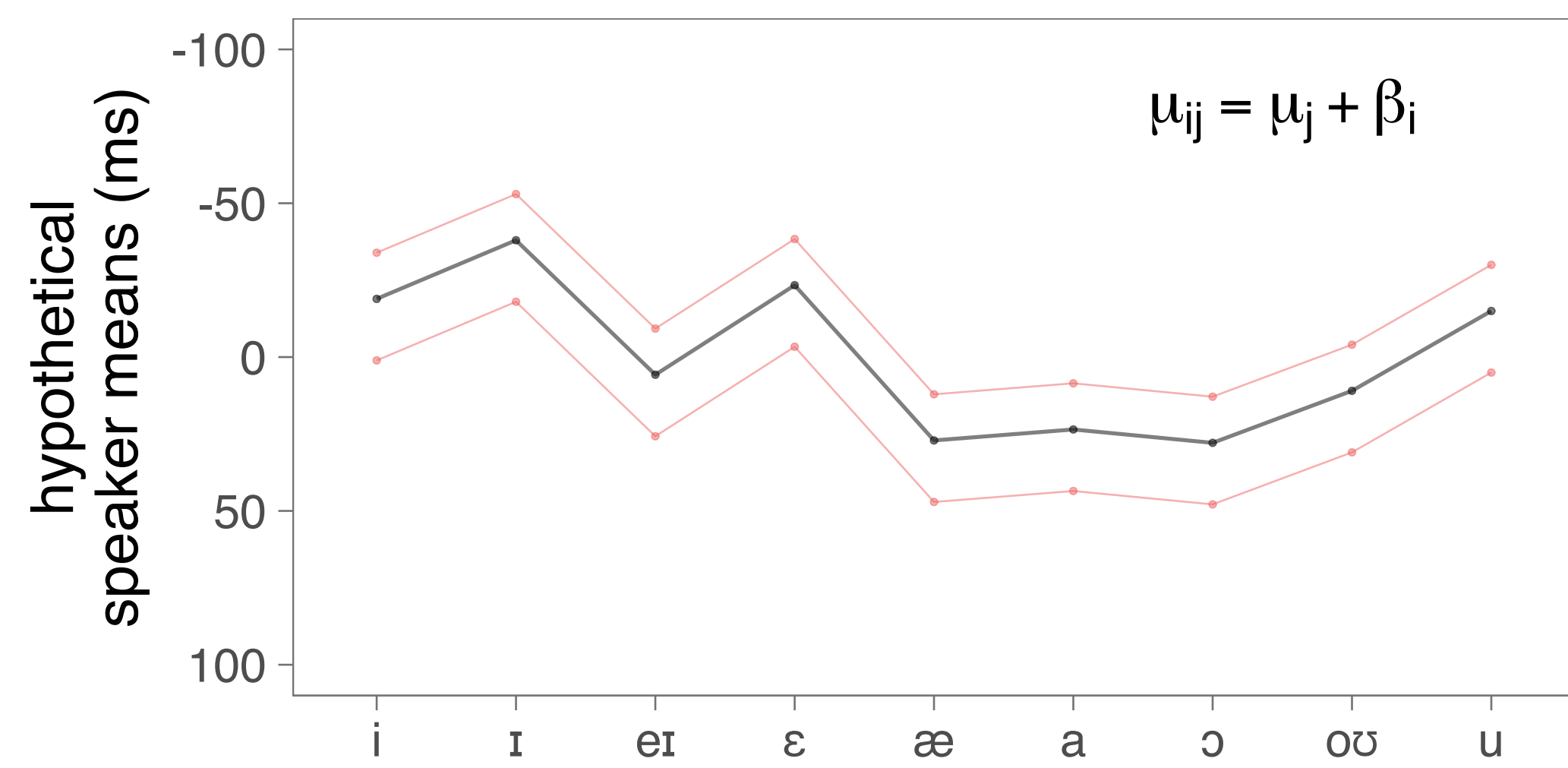
Vowel duration commonly varies across vowels of different qualities (e.g., House & Fairbanks 1952; Peterson & Lehiste 1960; Delattre 1962; Lindblom 1967; Lehiste 1970; Klatt 1973, 1976; Lisker 1973; Catford 1977; Keating 1985; Crystal & House, 1988; Hillenbrand et al. 1995)

- lower vowels > higher vowels
- tense vowels > lax vowels

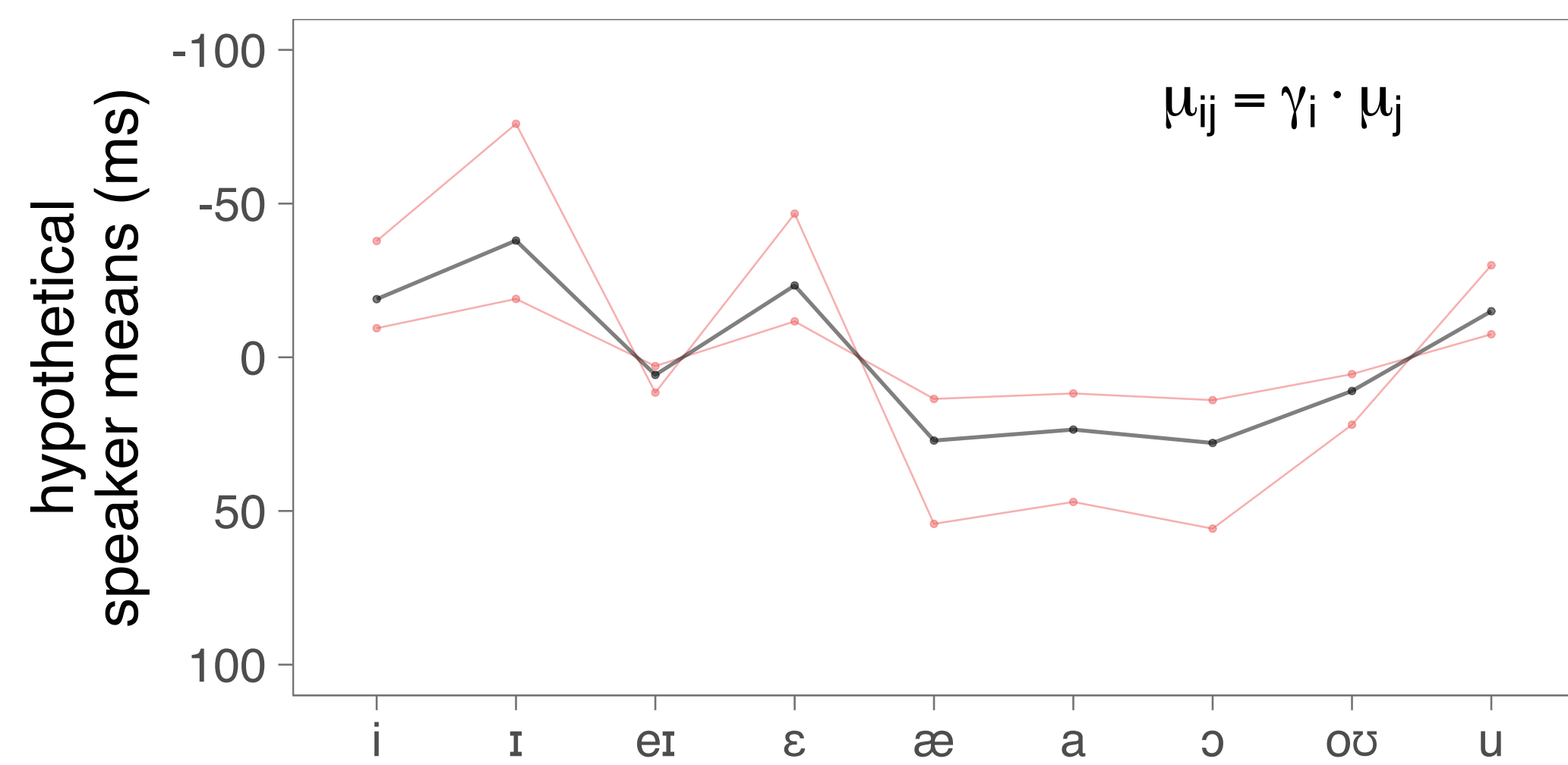
How does the **pattern of intrinsic vowel duration** differ across speakers of the same language? and what aspects of the pattern are (more) invariant?

Each speaker (red) could potentially differ from the population avg (black) by ...

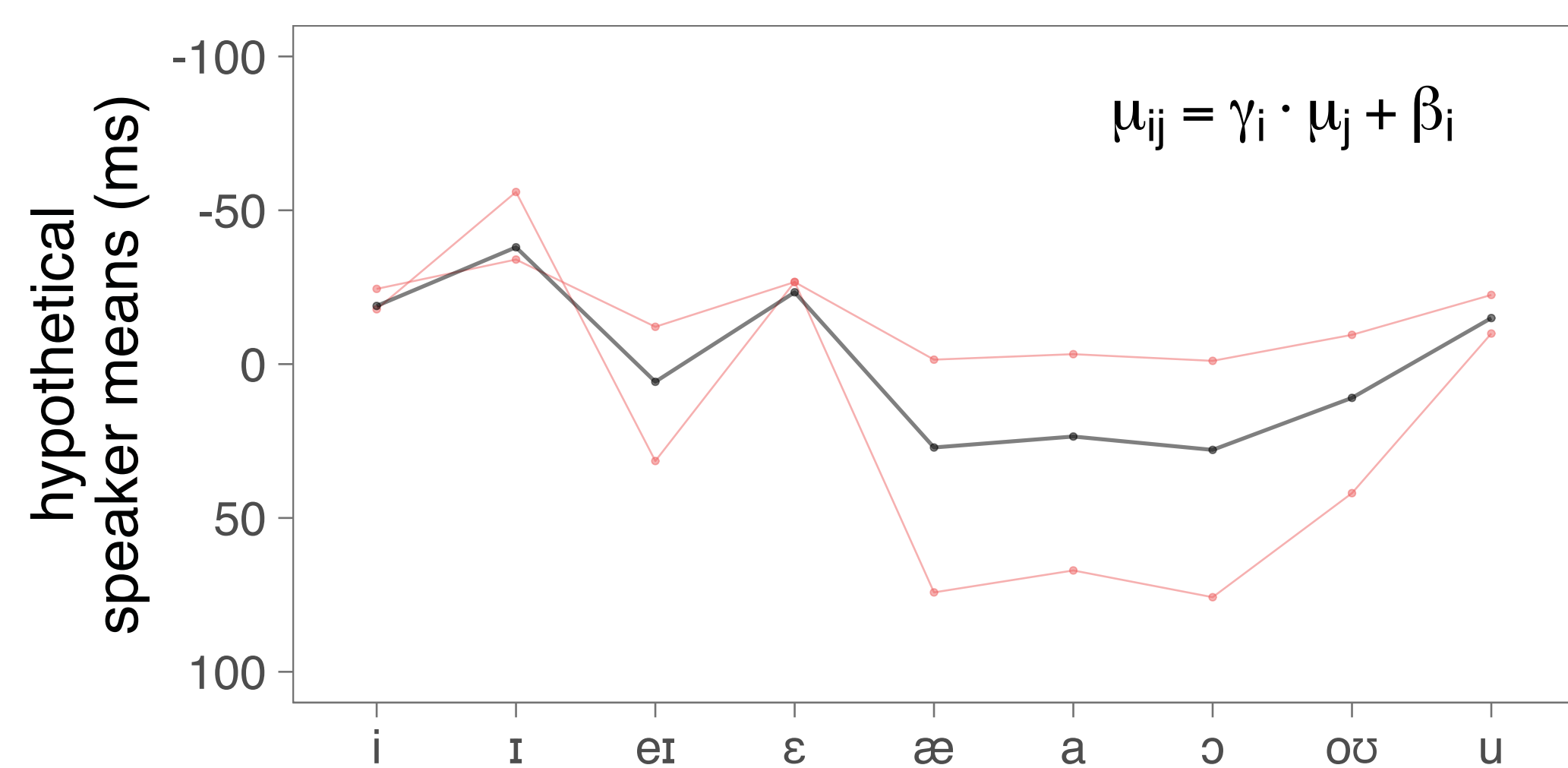
- **Uniform translation:** all vowels shorter/longer by same constant



- **Uniform scaling:** all durations compressed/expanded by same factor (scaling is equivalent to translation on a log scale, preserving ratios)



- **Uniform translation and scaling**

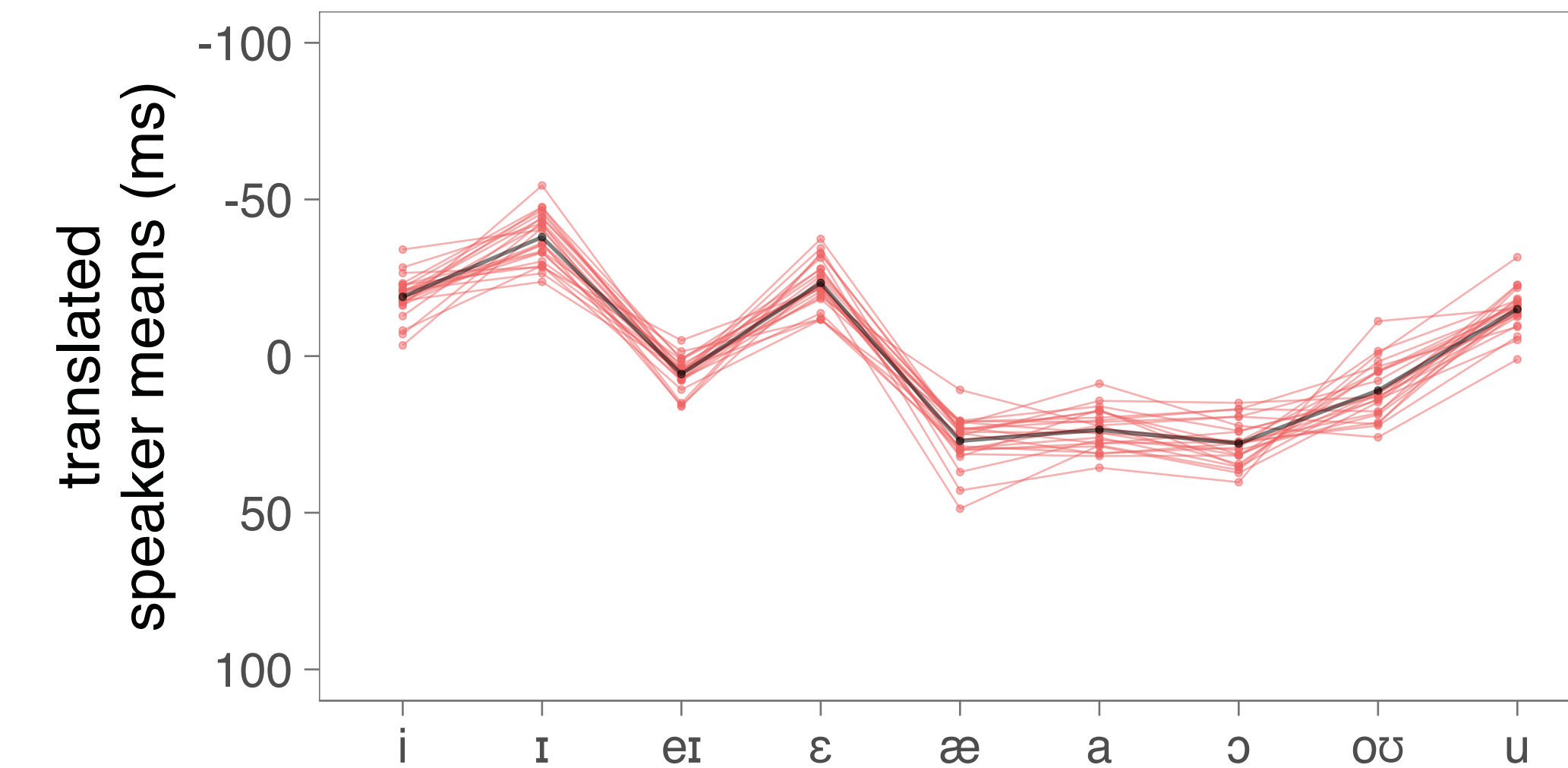
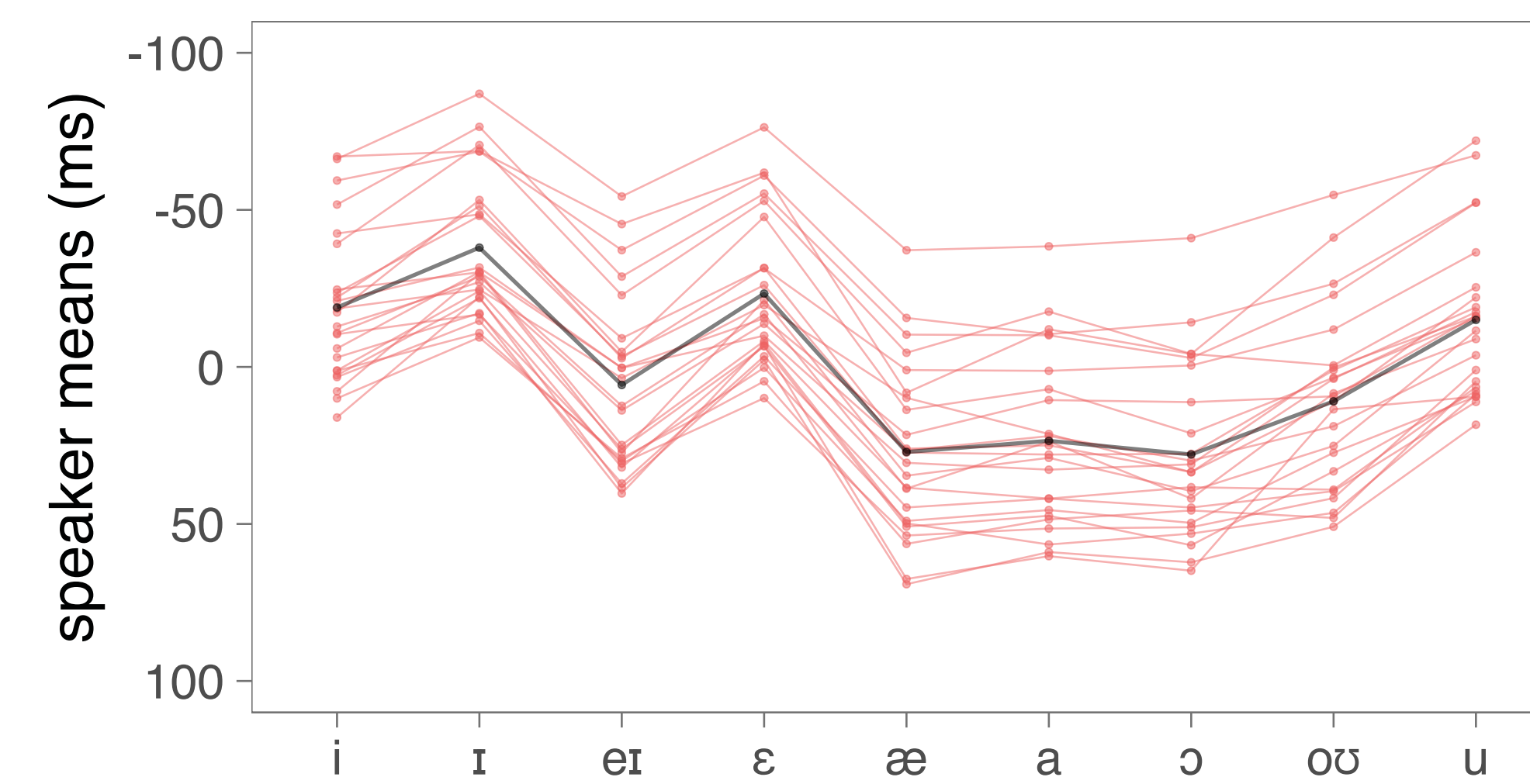


- **Non-uniform** effects of many conceivable types (e.g., selective shortening of low vowels, separate scaling factors for tense and lax vowels, etc.)

Case studies

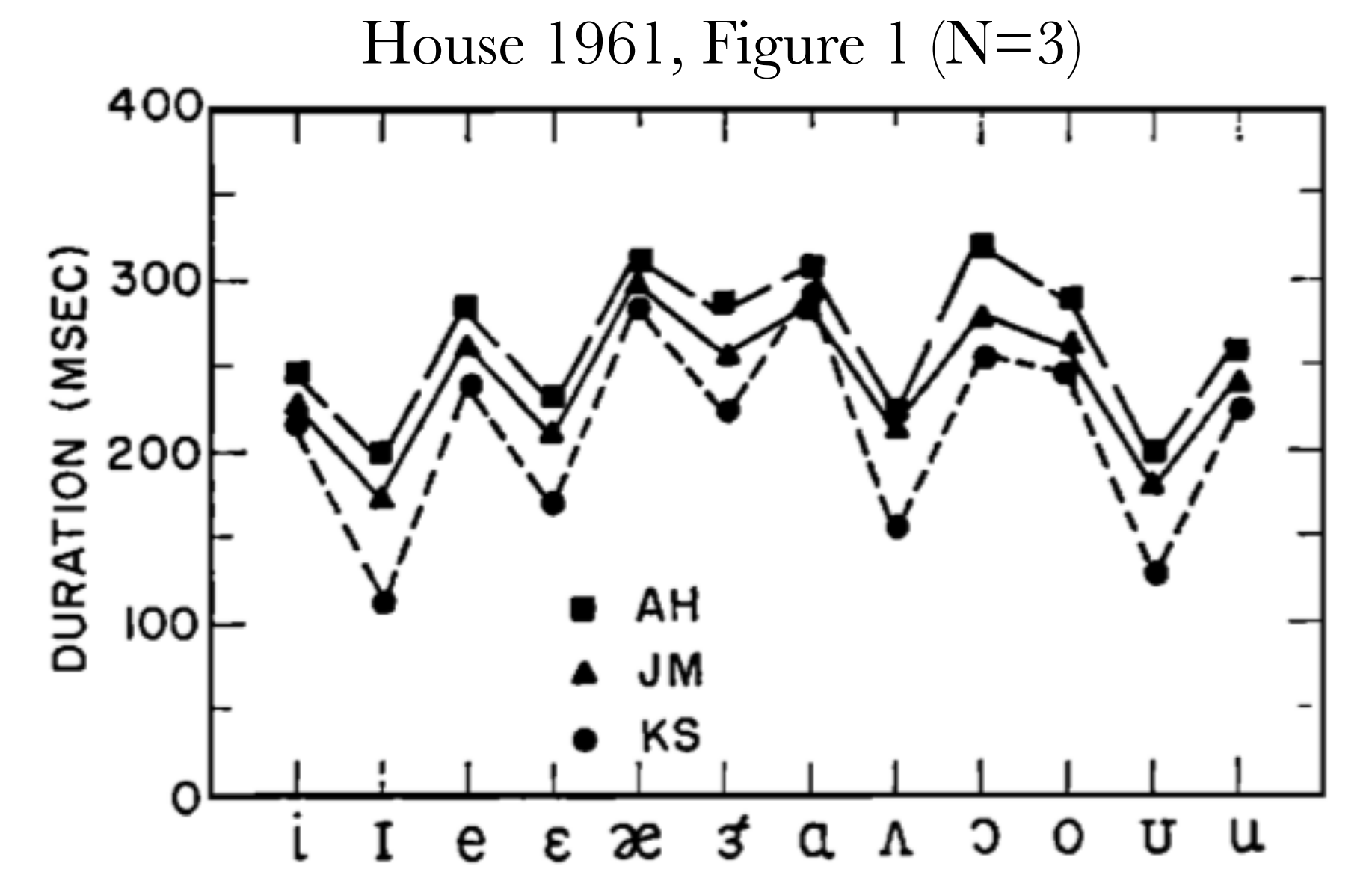
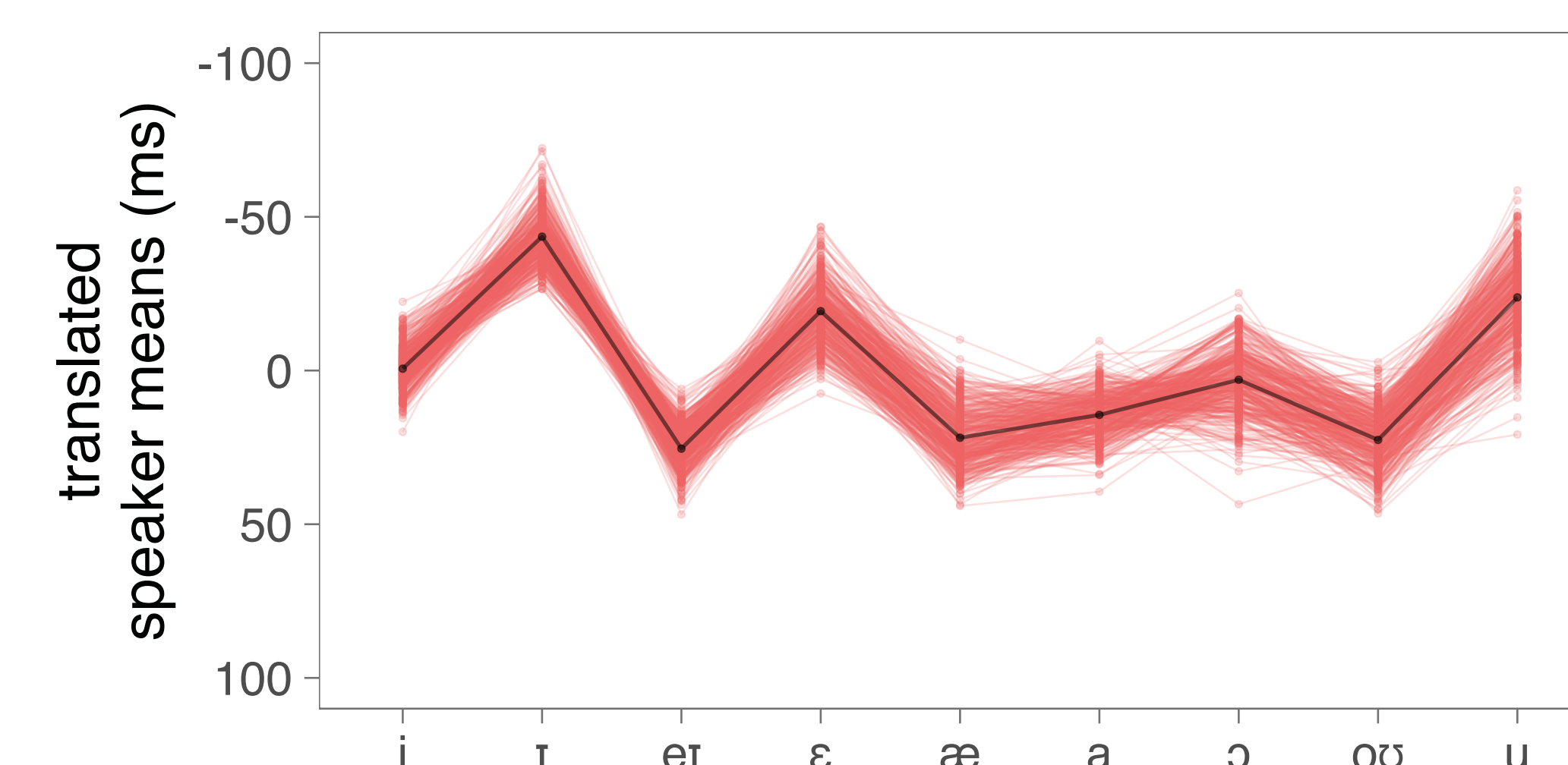
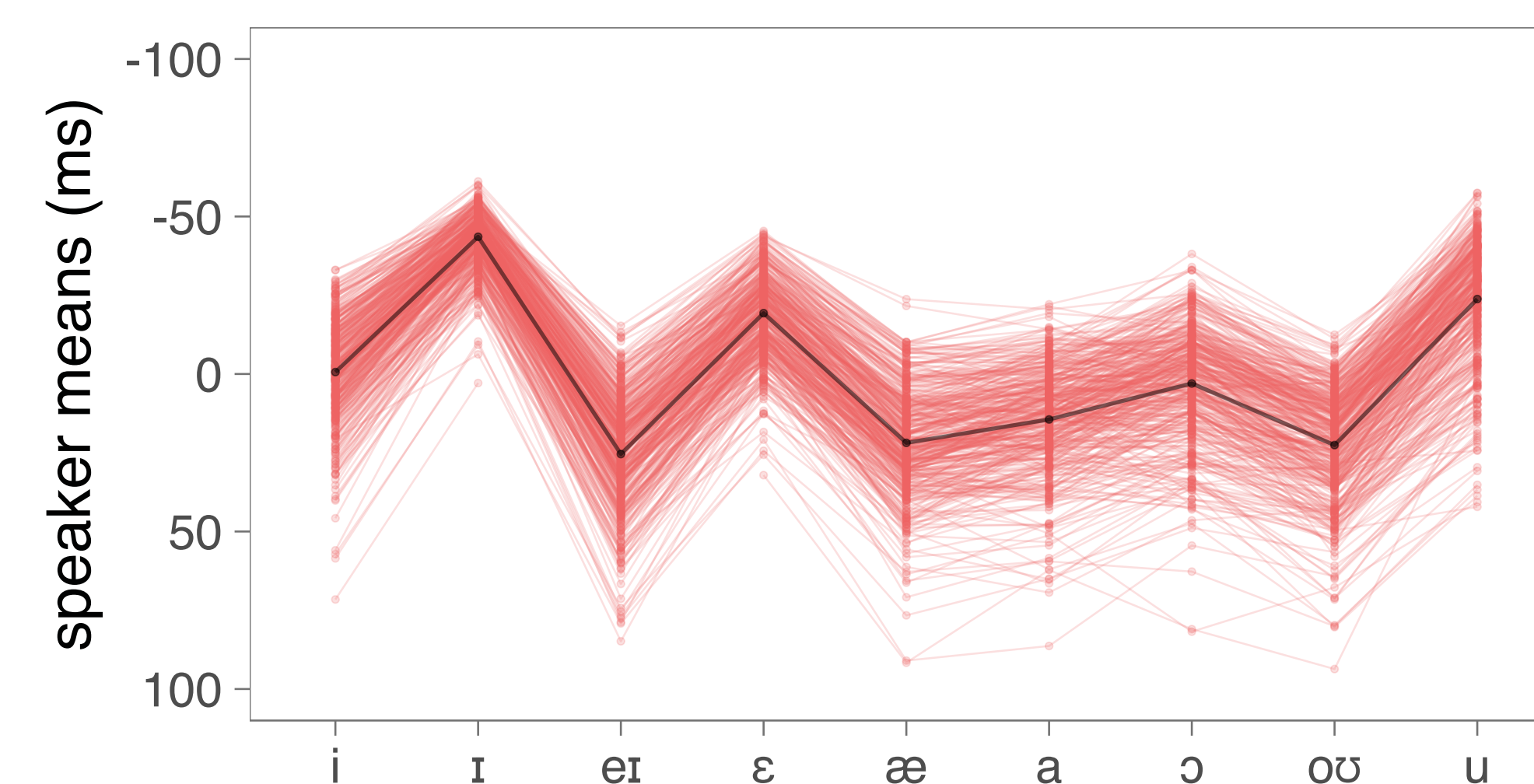
Isolated speech (recordings from Chodroff & Wilson 2014, *JASA*)

- American English speakers (N=24) produced [CVt] syllables (5 repetitions each) in carrier phrase
C = [p^h b t^h d k^h g] V = [i I e ε æ ɜ a ɔ oʊ u]
- Vowel tokens hand-segmented and **mean duration** of each vowel type calculated within speaker (subsequent to removing disfluencies and trimming bottom and top 1% of durations)
- Speaker-specific values centered by subtracting grand mean calculated over all vowel types



Connected speech (Mixer 6 Corpus, LDC2013S03; Brandschain et al. 2010; Chodroff et al. 2016)

- American English speakers (N=391, 209 female) read a common list of sentences from the Switchboard corpus
- Vowel boundaries identified with the Penn Phonetics Lab Forced Aligner from partially audited transcripts
- Speaker-specific mean duration for each vowel type was calculated as in the analysis of isolated speech (subsequent to removing stopwords and trimming bottom and top 1% of durations)



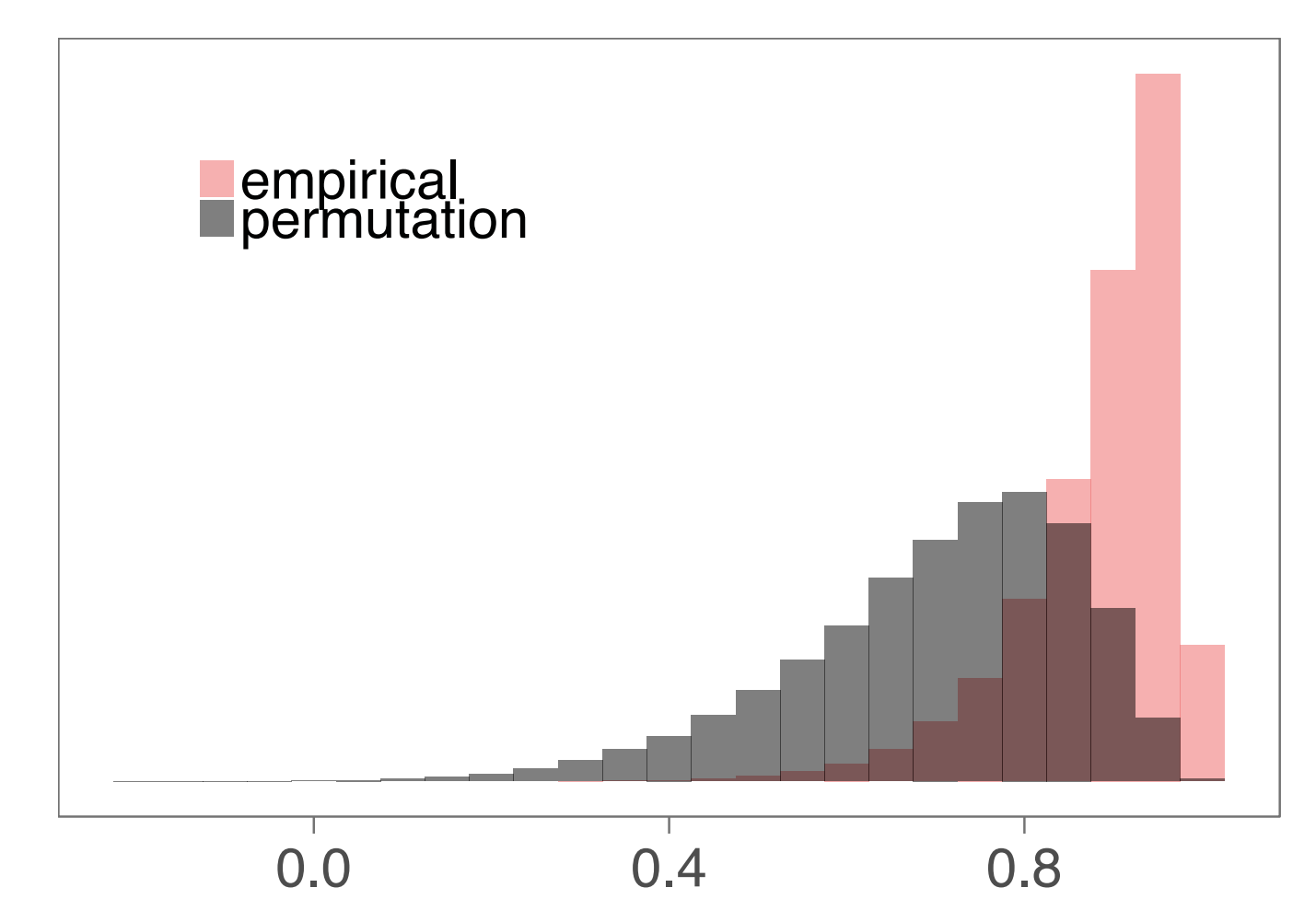
Uniform translation alone substantially reduces the differences among speakers

RMSE	Isolated	Connected
population avg	25.45	16.44
translation	6.99	8.94
scaling	25.40	16.16
translation & scaling	5.92	8.32

Patterns of intrinsic vowel duration are **highly correlated** across speaker pairs

Isolated: *median* .95, 95% within [.83, .99]
Connected: *median* .90, 95% within [.65, .98]

Pairwise speaker correlations



Discussion

- Intrinsic vowel durations are partly **motivated** by physical factors (e.g., jaw displacement for low vs. high vowels) but are nevertheless **controlled** by speakers (e.g., Westbury & Keating 1980; Solé & Ohala 2010) “However, if vowel duration is a controllable parameter, it is in principle available for language-specific [and speaker-specific, CW&EC] manipulation.” — Keating (1985:120)
- Near-isomorphism of vowel duration patterns across speakers of American English indicates that, at least within a broad speech community, individual-level control of this phonetic property is highly **restricted**
- Constraints on phonetic variation could arise from a number of sources, including usefulness of duration as a perceptual cue to vowel contrasts (e.g., Daniloff et al. 1968; Ainsworth 1972; Hillenbrand et al. 2000)
- Similar findings for stop VOT (Chodroff & Wilson 2017) and fricative place (Chodroff 2017) highlight the need for principles that explain **controlled but constrained** aspects of individual phonetic systems

References (selected)

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *J. Phon.*, 61, 30-47.
House, A. S. (1961). On vowel duration in English. *JASA*, 33(9), 1174-1178.
Keating, P. (1984). Universal phonetics and the organisation of grammars. *Phonetic Linguistics*, ed. V. Fromkin, Academic Press, 115-132.
Lehiste, I. *Suprasegmentals*. (1970) Cambridge, Mass.: MIT Press
Westbury, J. R., & Keating, P. A. (1980). Central representation of vowel duration. *JASA*, 67(S1), S37-S37.