Simon Fraser University
Language and Brain Lab

Pardo, J. S. (2006, April). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pardo, J. S. (2013, August). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*, 1-5.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017, November). Phonetic convergence across multiple measures and model talkers. *Attention, Perception & Psychophysics*, *79*, 637-659.

Keywords: phonetic convergence, speech production, speech perception, imitation, conversation

Phonetic Convergence

- **Pardo (2006):** *an increase in segmental and suprasegmental similarity of the speech of one talker to another*
  - **Pardo (2013):** *an increase in the similarity of acoustic-phonetic form between talkers* - has been demonstrated both when a talker listens passively to speech, and while talkers engage in social interaction
- **(Giles, Coupland, and Coupland, 1991):** communication accommodation theory - *talkers may have varied in the degree of convergence depending on conversational role* (ie. instruction giver is more dominant, greater convergence on the part of the receiver)
- Previous research showed that female talkers converged more than male talkers, and less dominant talkers converged more than more dominant talkers
- **Labov (1976, 1984):** *dialect formation & change result largely from opportunities for direct social contact among talkers and are influenced by social relationships between interacting talkers*
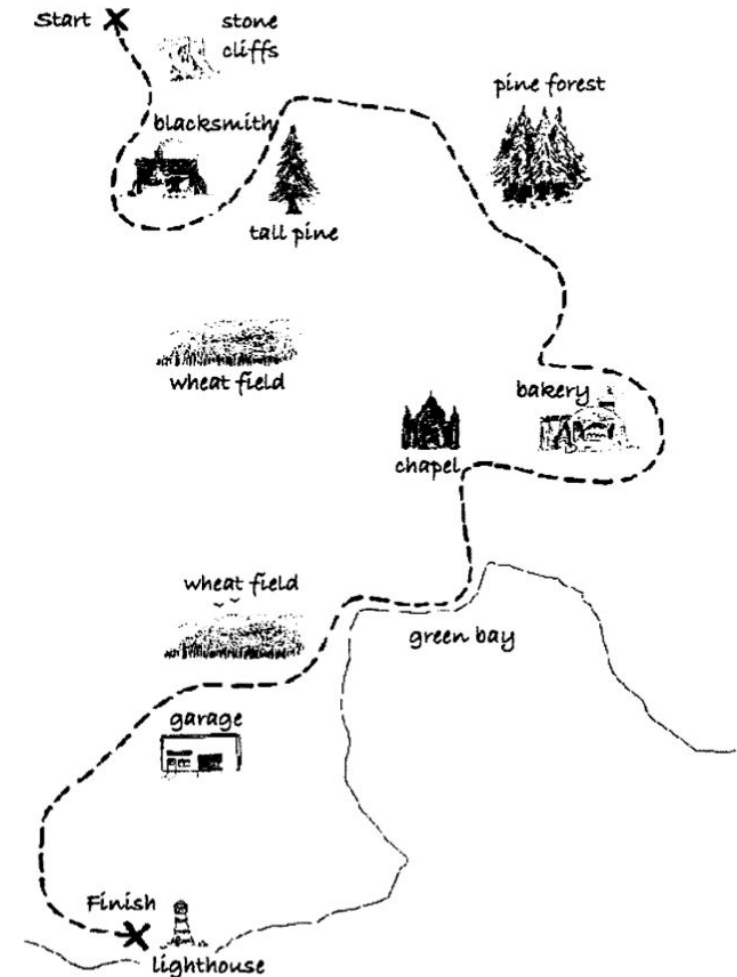
Pardo (2006)

- **Research Question:** What is the degree to which interacting talkers increase similarity in phonetic repertoire during conversational interaction?
- **Goal:** to determine if pairs of talkers *converged in phonetic repertoire over the course of a single conversational interaction*
  - Differs from past research in 2 ways:
    - (1) AXB test targeted: pronunciation of consonants and vowels as compared to sample items
    - (2) past studies used items produced after a delay of a few seconds from the sample presentation, Pardo (2006) used items produced immediately after the conversational interaction - not directly prompted by another talker's utterance

Pardo (2006) - Map Task

- **Map task:** 30-60 min
- 6 men & 6 women
- Participants blocked from seeing each other
- **Lexical items:** pre & post task carrier phrases "Number_is_the_;" and "Say_again" and landmark labels
- **Conversational Roles:** instruction giver, instruction receiver Enables assessment of phonetic convergence in an AXB perceptual test
- Same sex pairs, average F0 measured & pairs based on proximal F0

APPENDIX A: SAMPLE MAP TASK MAP FOR A GIVER

Pardo (2006) - Listening Test

- Listening test included 4 items:
  - ***abandoned monastery, green bay, walled city, wheat field***
- 30 undergrad participants
  - 2 items from early in conversation (before halfway point)
  - 2 items from later in conversation (after halfway point)
  - 3 repetitions of the same landmark label phrase, varied across trials
  - Equally considered giver-receiver, receiver-giver
- For each triplet, listener decided as quickly as possible which item (1st or last) sounded more similar to middle item (pronunciation)
  - ie. press "1" for 1st item, "0" for last item on keyboard
  - each trial presented 3 times in mixed random order
  - effects of timing, persistence/comparison type, talker role, pair sex tested within subject

Pardo (2006) - Convergence Assessment

- Evidence for phonetic convergence? *"...a talker's speech became more similar in pronunciation to the partner speech than it was before the interaction"*
  - based on landmark label phrases
- Pardo: time course of convergence - measured early & late in the conversational session, and in the post-task session
- Accuracy of map task assessed (effective communication)

AXB similarity test: pre & post task sessions compared to task repetitions
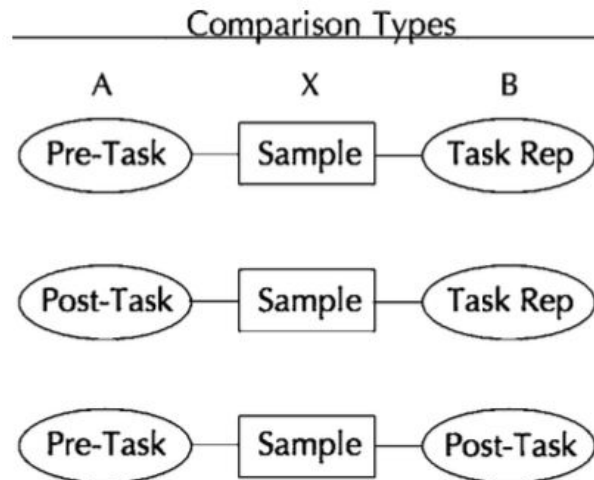
Pardo (2006) - Results

- Map task: 85% accuracy of route duplication overall, no significant differences in performances
- Repeated measures ANOVA
  - tested for effects of timing, persistence/comparison type, talker role, pair sex

TABLE I. Interaction between comparison type and role.

	R to GX	G to RX
Pretask vs task[a]	62	68
Post-task vs task	56	57
Pretask vs post-task[a]	58	65

[a]95% confidence intervals verified that means comparisons differed across these two rows and all measures were different from chance.
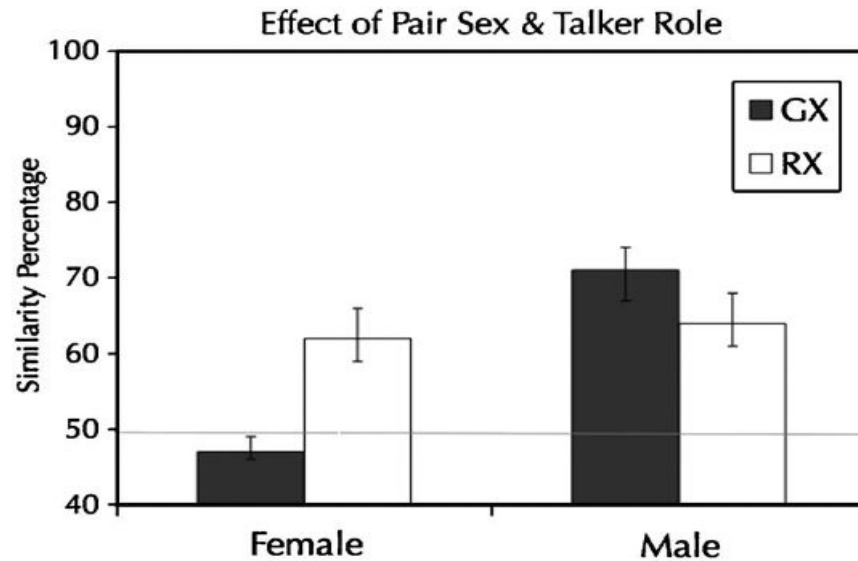
Pardo (2006) – Results

FIG. 2. AXB test interaction between *pair sex* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.
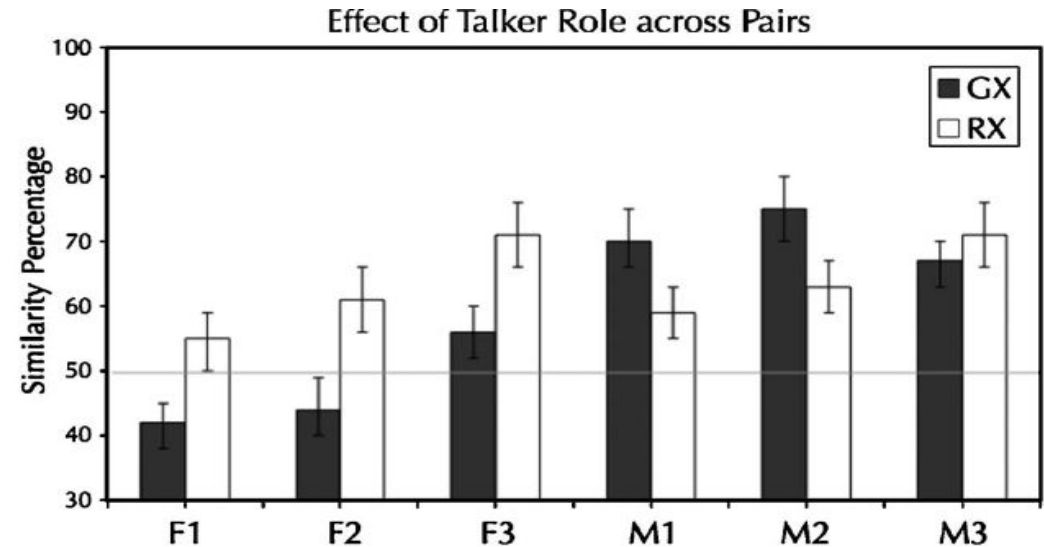

FIG. 3. AXB test interaction between *pair* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.

Pardo (2013)

- Previous research: Talkers converge to an interlocutor / auditory model on multiple acoustic-phonetic dimensions
- **Putman & Street (1984):** convergence among same-sex interviewers/interviewees considering speaking rate, turn durations, response rates - convergence only found in speaker rate
- **Sancier & Fowler (1997)**: long-term "gestural drift" - talkers vary acoustic-phonetic repertoire to sound more similar to a model talker/linguistic environment
- **Goldinger (1998):** talkers provided baseline & shadowed versions of mono- and bi-syllabic words
    - AXB task
    - established phonetic imitation in speech shadowing
    - replicated & extended by:
        - Namy et al. (2002) - influence of talker sex
        - Shockley et al. (2004) - variation in voice onset timing
- Pardo (2006)

Pardo (2013)

- Pardo notes the importance of a perceptual task & multiple attributes for 2 reasons:
  - (1) *"...not possible to characterize the phenomenon in a comprehensive manner without including multiple dimensions"*
    - ie. convergence in F0 in one item does not imply that the talker will always or only converge on F0
  - (2) social function of phonetic convergence
- Must consider:
  - collecting data from multiple listeners per talker pair
  - incorporate all levels of variability in analyses
  - include both acoustic & perceptual measures

Pardo (2013)

- Recent studies have examined single attributes:
  - Duration
  - Speaking rate
  - F0 / intonation contour
  - Intensity
  - Voice quality
  - Vowel spectra
  - Voice Onset Time
  - Lip aperture
  - Individual phonemic variants
- *Currently no compelling rationale/standard for choosing one acoustic attribute over another - a reliance on acoustic measures yields datasets that are *relatively inconsistent and chaotic*

Pardo (2017)

- A survey of 3 dozen reports & 1 new study
  - Overall, shadowing studies used very few model talkers & did not asses if phonetic convergence varied across same and mixed-sex pairings
  - Some studies showed effects of talker sex / word frequency, but other studies have failed to replicate these effects or report opposing patterns
- Phonetic detail in episodic memory: *talker-related details affect speech perception, persist in memory, and could support convergence in production* (Goldinger, 1998 for more details)

Pardo (2017)

- Table 1: only studies using shadowing / exposure tasks (not studies with conversational tasks)
- Measure include: AXB, f1 & f2 vowels, duration, F0, lip aperture, intensity, voice onset time, etc.
- More studies involved multisyllabic items than monosyllabic items
- Not all studies were equal in terms of mode or shadower sex

Table 1   Summary of noninteractive shadowing/exposure studies of phonetic convergence

Year	Authors	Models		Shadowers		Items		Measures
		Female	Male	Female	Male	Mono	Multi	
1998	Goldinger	5	5	12	12	80	80	AXB
2002	Namy et al.	2	2	8	8		20 LF	AXB
2003	Nye & Fowler, Exp. 2		2	?/2	?/2		12	AXB
2004	Goldinger & Azuma (expo)	2	2	6	6		160	AXB
2004	Shockley et al.	1	1	12	12		80 LF	AXB, VOT ptk
2010	Miller et al.	1	1	8	8		74 LF	AXB
2012	Babel & Bulatov		1	12	7	15	24	AXB, F0
2013	Babel et al.		1	33	8	18		AXB, vowel
2013	Miller et al.	2	2	8	8		74 LF	AXB
2013	Pardo, Jordan, et al.	10	10	10	10	80		AXB, vowel F1F2, Dur, F0
2014	Babel et al.	4	4	10	10	15 LF		AXB
2015	Walker & Campbell-Kibler	4		36		70		AXB, vowel F1F2, rF3
2016	Dias & Rosenblum	1		32 & 24			120	AXB
2004	Vallabha & Tuller		3		3	Vs		vowel F1F2
2007	Delvaux & Soquet (expo)	5		12		3	1	vowel F1–F3
2007	Gentilucci & Bernardis, Exp. 2	2	2	10			/aba/	vowel F1F2, lip aperture, F0, dur, intensity
2009	Tilsen		1	6	6	Vs		vowel F1F2
2010	Babel		1	34	8	25		vowel F1F2 DID
2012	Babel		2	60	51	50 LF		vowel F1F2 DID
2012	Nguyen et al.		1	33	9	40		vowel F1 o aw
2013	Dufour & Nguyen	1		16	4		66	vowel F1 eE French
2003	Fowler et al., Exp. 4	1		?/24	?/24		48 VCVs	VOT ptk
2010	Sanchez et al.	1		35		6 CVs		VOT p
2011	Abrego-Collier et al. (expo)		1	?/48	?/48	17	55	VOT ptk
2011	Nielsen, Exp. 1 (expo)		1	?/27	?/27	?/120	?/120	VOT pk
2013	Olmstead et al.	1		20	12	11 CVs		VOT b-p
2013	Yu et al. (expo)		1	?/84	?/84	17	55	VOT ptk
2013	Garnier et al.	1	1	4	11	Vs		F0
2013	Mantell & Pfordresher	1	1	69	86		12	F0
2013	Postma-Nilsenová & Postma	2	2	67	21		16	F0
2013	Sato et al.	3	3	30	30	Vs		F0, F1 ieE
2013	Wisniewski et al.	1	1	8	8		12	F0
2008	Mitterer & Ernestus	1		?/18	?/18		28 CVVC	phonemic/r/allophones Dutch
2011	Honorof et al.		1	?/37	?/37		4 VCVs	phonemic/l/allophones English
2013	Mitterer & Müsseler	1		9	3		100	phonemic allophone pairs German

Studies in the table are grouped according to the measures used to assess phonetic convergence (last column)

Pardo (2017) - Method

- Talkers: 108 (54 female), native English speakers
  - Split into 2 groups: (1) provided model utterances (2) provided baseline & shadowed utterances in random same- and mixed-sex pairings with model talkers
  - Model utterances:12 talkers provided model recordings of all 160 words in 3 randomized blocks
  - Shadower utterances: 92 talkers provided baseline & shadowed recordings of the word set in 32 same-sex famel, 30 same-sex male, 30 mixed-sex pairings
- Listeners: 736, native English speakers
  - provided holistic pronunciation judgements in AXB perceptual similarity tests
  - 3 repetitions of the same lexical items with presented, models items as X, shadowers baseline & shadowed versions of the same item as A and B, counterbalanced
  - 184 separate tests of 80 words each, each word triad was presented 4 times: once in each order in 2 randomized blocks
  - 4 listeners per shadower test: monosyllabic & bisyllabic

Pardo (2017) - Acoustic Measures

- Focused on monosyllabic words
  - balanced vowel identity &
  other segmental characteristics
- Vocalic duration, F0, vowel
  formants (F1 & F2) at midpoint
- Mixed-effects regression
  analysis: talker sex, lexical
  factors, acoustic convergence on
  AXB perceptual similarity

Appendix A: Word sets

Bisyllabic		Monosyllabic	
Low Frequency	High Frequency	Low Frequency	High Frequency
active	basis	babe	bad
balance	become	bathe	bag
beacon	before	beak	beach
bicep	better	bean	beam
captain	between	boot	beat
career	beyond	cage	bet
careful	city	cake	bone
cavern	common	cop	check
coffee	country	cot	death
cousin	father	dab	dock
deport	figure	dad	foot

Sample of word sets used

Pardo (2017) - Results

- Overall AXB phonetic convergence proportion averaged .56 (chance = .50)
  - Shadowers converged to model talkers, observed effect was characteristically subtle & compared to those in previous shadowing studies
- Overall, convergence was equivalent across female & male shadowers and male talkers
  - lack of reliable effects of talker sex challenges the idea that female talkers converge more than males
- Phonetic convergence influenced by word type - bisyllabic words evoked greater convergence than monosyllabic words (.57 > .55)
- Acoustic attributes:
  - Duration, F0, vowel spectra converted into difference-in-distance (DID) scores; compared baseline differences bw each shadower (DID = baseline distance - shadowed distance)
  - Values greater than - = acoustic convergence

Pardo (2017) - Results

- Female shadowers more susceptible to lexical effects
  - (Top) Female shadowers converged more to low-frequency words, male talkers not affected by word frequency
  - (Bottom) Female shadowers showed a marginally stronger difference in convergence to mono vs. bisyllabic words
  - No interactions bw model sex & lexical factors in AXB convergence
- Impt to note that the 3 studies that reported greater convergence of female shadowers all used low frequency words (Babel, 2014; Namy, 2002; Miller, 2010)
- Word type influenced phonetic convergence, shadowers converged more to bisyllabic than monosyllabic words
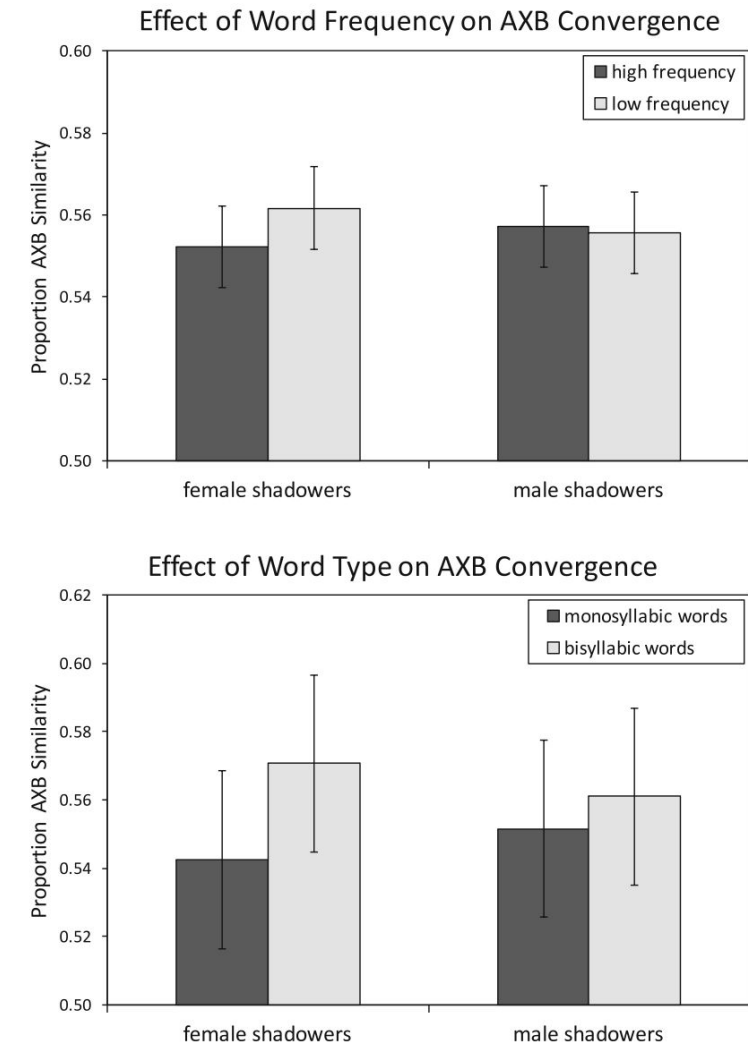

Fig. 2

Takeaways for our experiment

- Acoustic features - which to focus on?
- Mono/multi/bisyllabic targets?
- Shadow vs. conversational task
- Details of conversational pairs (ie. gender, F0)
- Perceptual task
- Time-course of convergence
- Map task? What other ways to elicit "natural" conversation?
- Roles in conversation
- Word frequency
- Measures
- Adverse listening conditions
- Naturalness of map task?