

Our Project

Acoustic and lexical effects on speech perception in Kaqchikel (Mayan)

LSA 2017

Ryan Bennett¹, Kevin Tang¹, Juan Ajsivinac²

kevin.tang@yale.edu & ryan.bennett@yale.edu

¹Yale University

²Independent scholar

Jan 5th–8th, 2017

Our project: the production-perception-lexicon interface in Kaqchikel (Mayan).

Methodological challenge: to model the **production and perception** of an **under-resourced and under-studied** language with **small and noisy** data collected in the **field**.

Outline

Goals of the talk:

- ▶ Report on:
 - ▶ Construction of spoken and written corpora.
 - ▶ An AX discrimination study on the perception of stop consonants.
- ▶ Examine:
 - ▶ The effect of acoustic and lexical factors on speech perception.
 - ▶ The time course of such effects.

Outline

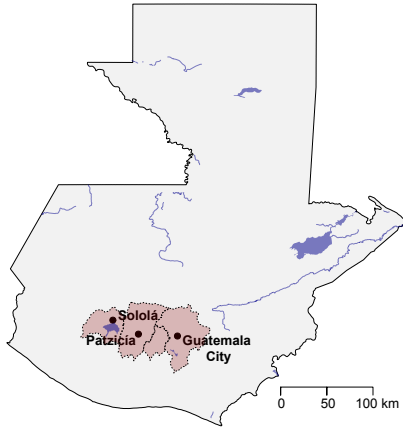
General findings:

- ▶ Both acoustic and lexical factors affect speech perception in Kaqchikel.
 - ▶ Indirect validation of small corpora for speech perception research.
- ▶ Both acoustic and lexical factors kick in early, and decay over time.
- ▶ Rich, experience-based factors influence perception even in low-level tasks which do not require lexical access.

Kaqchikel

Kaqchikel is a K'ichean-branch Mayan language spoken in the central highlands of Guatemala (over 500,000 speakers, Richards 2003, Fischer &

R. M. Brown 1996: fn.3).



Phonemic consonants

	Bilabial	Dental/ alveolar	Post-alveolar	Velar	Uvular	Glottal
Stop	p b	t tʔ		k kʔ	q qʔ	ʔ
Affricate		ts tsʔ	tʃ tʃʔ			
Fricative		s	ʃ	x ~ χ		
Nasal	m	n				
Semivowel	w		j			
Liquid		l r				

(Campbell 1977, Chacach Cutzal 1990, Cojtí Macario & Lopez 1990, García Matzar et al. 1999, Majzul et al. 2000, R. M. Brown et al. 2010, Bennett 2016, etc.)

Perception study: procedure

Kaqchikel speakers heard pairs of [CV] (onset) or [VC] (coda) syllables.

- ▶ Vowels were always identical, but consonants could be different.
- ▶ Items embedded in speech-shaped noise generated from spoken corpus (0dB SNR, after amplitude normalization; LTAS over 4 hours of corpus).

Participants asked to respond SAME or DIFFERENT on a button box.

- ▶ **Assumption:** incorrect SAME responses indicate perceptual similarity between [C₁]~[C₂] pairs.

Perception study: stimuli

Item properties:

- ▶ V ∈ /a i u/
- ▶ C ∈ all consonants of Kaqchikel
 - ▶ Target pairs: C ∈ /p b t tʔ k kʔ q qʔ (?)/ (no affricates)
 - ▶ Filler pairs: any other consonant combination
- ▶ Syllables recorded by native speaker of Patzicía Kaqchikel (Ajsivinac).

Each participant heard 200 total trials (6000 pairs, in 30 randomized lists).

Perception study: presentation

Timing details:

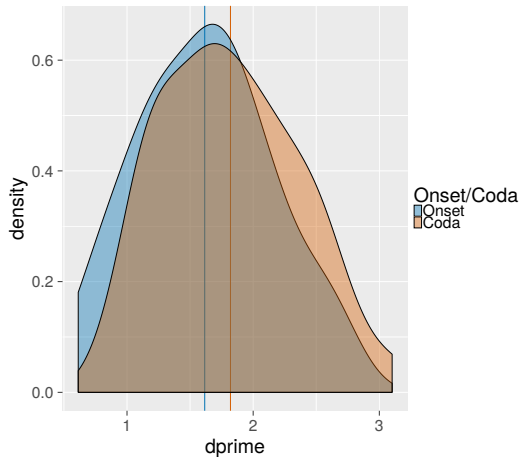
- ▶ ISI = 800ms (250ms of noise padding before/after each syllable + 300ms silence between items)
- ▶ Inter-trial interval = 1500ms
- ▶ Up to 10 seconds to respond without receiving a warning.
 - ▶ Most responses under 1 sec. (mean RT = 854ms, median RT = 664ms)

Moderate ISI and response times may have favored a **linguistic mode of speech processing**.

(Pisoni 1973, 1975, Pisoni & Tash 1974, Fox 1984, Werker & Logan 1985, Kingston 2005, Babel & Johnson 2010, McGuire 2010, Kingston et al. 2016 and references there)

General findings

Relatively good discrimination: $d'_{\mu} \approx 1.75$



Perception study

45 participants (44 completed the study).

- ▶ All speakers of Patzicía Kaqchikel.
- ▶ Good mix of ages and genders.
 - ▶ 13 male, 31 female
 - ▶ Ages 18-50 (mean = 26, median = 25, SD = 6.2)

General findings

Dorsals confusable with each other, apart from /k^ʔ/ (see also Shosted 2009).

- ▶ Onset [TV] d' : /k q q^ʔ/~ /k q q^ʔ/ 1.23 < all others 1.65
- ▶ Coda [VT] d' : /k q q^ʔ/~ /k q q^ʔ/ 1.50 < all others 1.85

/β/ frequently confused with /p β^ʔ/.

- ▶ Onset [TV] d' : /β/~ /p β^ʔ/ 0.77 < /β/~all others 1.61; highest d' rank = 32/36
- ▶ Coda [VT] d' : /β/~ /p β^ʔ/ 1.16 < /β/~all others 1.88; highest d' rank = 31/36

Corpus criticism

- ▶ Spontaneous speech is *naturalistic*, but. . .
- ▶ . . . leads to data sparsity (cf. Xu 2010)
 - ▶ /t^ʰ/ is rare (18, <1% of stops; England 2001, Bennett 2016)
 - ▶ Large skew toward prevocalic [CV] stops (>85%)
- ▶ Narratives, not dialogues (cf. CALLHOME, Switchboard)

Corpus criticism

- ▶ Not huge — poor estimates of low frequency words (Brysaert & New 2009)
- ▶ Not terrifically speech-like — too religious and governmental.
- ▶ Noisy — OCR errors, typos, new-line hyphens. . .
 - ▶ Applied various filters to clean up the corpus (see Appendix).

Corpus construction

To test for an effect of lexical measures on speech perception, we compiled a text corpus of Kaqchikel:

- ▶ Corpus size: 1 million word tokens.
 - ▶ Constructed from existing religious texts, spoken transcripts, government documents, and educational books.
- ▶ Compare:
 - ▶ Kučera & Francis (1967): 1.014 million words of English
 - ▶ van Heuven et al. (2014): 201 million words of English

Acoustic similarity

Expectation: greater acoustic similarity predicts greater perceptual similarity.

Two kinds of acoustic similarity:

- ▶ STIMULUS SIMILARITY
- ▶ CATEGORY SIMILARITY: similarity of two phoneme categories based on *prior phonetic experience*.
 - ▶ Specifically: **category overlap**

Acoustic similarity

We used DYNAMIC TIME WARPING to estimate acoustic similarity

(Sakoe & Chiba 1971, Mielke 2012)

- ▶ Stimulus similarity: over stimulus pairs.
- ▶ Category similarity:
 - ▶ Over all possible [CV] and [VC] pairings in the acoustic corpus
 - ▶ Pairs matched for stress and vowel quality.

DTW gives us a similarity metric for each pair of stimuli/sounds.

Results

Analyzed participant accuracy with a mixed-effects logistic

regression in R (R Development Core Team 2013, Bates et al. 2011)

Parameters:

- ▶ Fixed effects:
 - ▶ All acoustic and lexical factors mentioned above (no interactions).
 - ▶ Response time (z-scored by participant)
- ▶ Random effects:
 - ▶ Participant
 - ▶ By-participant slopes for lexical factors
 - ▶ Nuisance factors (item, list, stimulus order, onset/coda)

Full model reduced by step-down model selection.

Lexical factors

Well-known that lexical factors interact with speech perception:

- ▶ Wordhood (e.g. Ganong 1980)
- ▶ Word frequency (e.g. C. R. Brown & Rubenstein 1961, Broadbent 1967, Vitevitch 2002, Felty et al. 2013, Tang & Nevins 2014, Tang 2015: Ch.4)
- ▶ Bigram frequency (e.g. Rice & Robinson 1975, Carreiras et al. 1993, Barber et al. 2004, Albright 2009, González-Alvarez & Palomar-García 2016)
- ▶ Segmental frequency (e.g. Kataoka & Johnson 2007, Tang 2015: Ch.4, Bundgaard-Nielsen et al. 2015)
- ▶ Neighborhood density (e.g. Luce 1986, Yarkoni et al. 2008, Bailey & Hahn 2001, Gahl & Strand 2016)
- ▶ Functional load/Presence of minimal pairs (e.g. Martinet 1952; Baese-Berk & Goldrick 2009, Graff 2012, Goldrick et al. 2013, Hall & Hume submitted)
- ▶ Etc.

Explanatory factors

	β	SE(β)	t	p-value
(Intercept)	0.8042	0.1621	4.963	6.95e-07***
Acoustic stimulus similarity	-1.0720	0.1151	9.316	2e-16***
Acoustic category similarity	-0.3876	0.1238	3.131	0.00174**
Functional load	0.4653	0.1649	2.822	0.00477**
Distributional overlap	-0.6320	0.1607	3.933	8.38e-05***
Word token frequency diff.	0.1848	0.1068	1.731	0.08353

Stimulus similarity and category similarity

Both stimulus similarity and category similarity had an effect on discriminability in the perception study.

Possible interpretation:

- ▶ Discrimination is mediated by some representation of prior phonetic experience.
- ▶ These representations include rich acoustic detail for individual phoneme categories.
- ▶ Consistent with exemplar-type theories of lexical representation
(e.g. Pierrehumbert 2001, 2016, Johnson 2005, Gahl & Yu 2006 and references there)

Time course

Assumption: segment-level phonetic processing occurs **prior to lexical activation** in speech processing.

(e.g. Fox 1984, Norris et al. 2000, Kingston 2005, Babel & Johnson 2010, Kingston et al. 2016, etc.)

Predictions about the time-course of effects:

- ▶ Acoustic factors > Lexical factors
- ▶ Segment-level > Word-level

Lexical Factors – Contrastiveness

Both functional load and distributional overlap play a role in discrimination.

A possible interpretation:

- ▶ Discrimination is mediated by **how contrastive** two phonemes are
 - ▶ Importance for minimal contrasts.
 - ▶ Relative predictability.
- ▶ The perceptual space is warped by contrastiveness.
 - ▶ Consistent with Hall's (2012) Probabilistic Phonological Relationship Model.

Time course effects

Responses binned according to by-participant RT terciles.

	Early ($\mu \approx 400\text{ms}$)	Middle ($\mu \approx 650\text{ms}$)	Late ($\mu \approx 1200\text{ms}$)
Acoustic stimulus similarity	-1.4515***	-1.1651***	-0.74647***
Acoustic category similarity	-0.6544**	-0.3020	-0.28756*
Functional load	0.9001**	0.4116	0.28513
Distributional overlap	-1.1437***	-0.8765***	-0.27972
Word token frequency diff.	0.2671 ^{n.s.}	0.2314 ^{n.s.}	0.06068 ^{n.s.}

Time course effects

Predictions about the time-course of effects:

- ▶ Acoustic factors > Lexical factors
- ▶ Segment-level > Word-level

Not borne out!

- ▶ Acoustic measures active early, and weaken over time.
- ▶ Same pattern for lexical measures (functional load, distributional overlap).
- ▶ Includes an experience-based measure of acoustic similarity (acoustic category distance)

Conclusions

Three caveats:

- ▶ Classic findings of late time course for lexical effects involve LEXICAL ACCESS (e.g. Ganong effect, Ganong 1980, Fox 1984, etc.)
 - ▶ Not clear that our 'lexical' measures—functional load, distributional overlap—involve lexical access in the same sense.
- ▶ Our ISIs may have been too long to 'catch' a purely pre-lexical stage of processing, even for fast response times (ISI = 800ms)
- ▶ Gradual decay (rather than increase) in strength of lexical effects over time may be more consistent with autonomous, feed-forward models (e.g. MERGE) than richly interactive models (e.g. TRACE) (TRACE, McClelland & Elman 1986, McClelland et al. 2006; MERGE, Norris et al. 2000; see again Kingston et al. 2016).

Conclusions

Our results suggest:

- ▶ Speech perception is mediated by phonetically rich memory traces associated with phonemic categories (exemplar theory).
- ▶ Lexical effects related to a graded notion of contrastiveness may affect speech perception.
- ▶ Lexical factors may have kicked earlier than predicted by 'modular' models of speech processing.
 - ▶ Did not find evidence that acoustic/phonetic processing precedes lexical activation.
 - ▶ Suggests co-activation of low-level and high-level factors. (McClelland & Elman 1986, McClelland et al. 1986, 2006)
 - ▶ Such activation appears to decay fairly quickly. (See too Kingston et al. 2016)

Conclusions

Small, noisy corpora can make valuable contributions to speech perception research — provided they are carefully processed.

References available on request.

Slides available for download at

http://tang-kevin.github.io/Files/Slides/Bennett_Tang_LSA2017.pdf