When to Hold and When to Fold:
Detecting Structural Changes in Statistical Learning

Benjamin D. Zinszer & Daniel J. Weiss, Pennsylvania State University

Background

Statistical Learning

A broadly-applied cognitive faculty for:
- Tracking regularities in input
- Learning complex relationships such as object naming, visual pattern recognition, acoustic stream segmentation

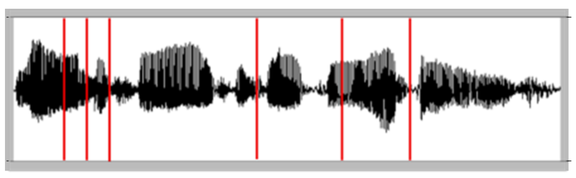
Widely applied in research of language acquisition, e.g.,
- speech segmentation (e.g., Saffran, Newport, & Aslin, 1996)
- word learning (e.g., Yu & Smith, 2007)
- phonetic learning (e.g., Maye, Werker, & Gerken, 2002)
- grammar learning (e.g., Gomez & Gerken,)

Input generated for artificial languages is typically invariant…
- Few studies have used multiple, incompatible structures
- Little is known about how learners contend with variability
  - Within a structure
  - **Between multiple structures**

How can learners infer the presence of multiple statistical structures? (Qian, Jaeger, & Aslin, 2012, *Frontiers in Psychology*)
- Variation in input can be interpreted in many ways (random deviations, change to current structure, new structure)
- **Inference about multiple structures from variation is analogous to Piaget's (1985) description of assimilation and accommodation**
- **Prerequisite to language acquisition in statistically noisy environment**

Speech Segmentation

A domain of statistical learning: *How can an acoustic stream be divided into discrete, meaningful units?*



Can pauses adequately mark the boundaries between words?

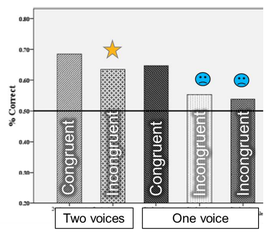
Adults and children can track transitional statistics between units to infer structure in a stream (Saffran, Newport, & Aslin, 1996)

lookatttheprettybaby
whataprettyflower
youareacutebaby

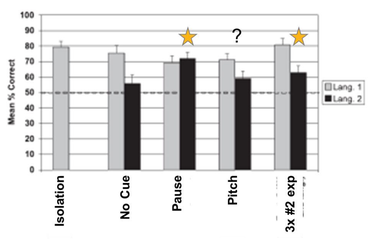
Segmentation of Two Structures (or Languages)

A majority of the world's population is bilingual (or multilingual), requiring them to learn two or more structures in language acquisition

Language learners segmenting speech *without* an overt cue to change in context (new structure):
- Can learners detect the presence of a new structure?
- Can learners track and learn *both* structures?



Interference Effect – Neither language is learned
(Weiss, Gerfen, & Mitchel, 2009, *Language Learning and Development*)



Primacy Effect – Only the first of two incongruent languages is learned
(Gebhart, Aslin, & Newport, 2009, *Cognitive Science*)

References

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33*, 1087-1116.
Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.
Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development (Vol. 985)*. Chicago: University of Chicago Press.
Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology, 3*(July), 228.
Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. Journal of *Memory and Language, 621*(35), 606-621.
Weiss, D. J., Gerfen, C., & Mitchel, A. (2009). Speech Segmentation in a Simulated Bilingual Environment: A Challenge for Statistical Learning? *Language Learning and Development, 5*(1), 30-49.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

We wish to thank Dr. Richard Aslin and Timothy Poepsel for comments and discussion, as well as Tzuyu Chang for assistance with data collection.

Research made possible through:
National Institutes of Health Grants to Daniel Weiss
NIH R01 HD067250-01

Question

What *implicit* cues to a new structure are available to learners?
- Studies above have differed in duration, number of exposures, switches between structures, degree of incongruity between the structures.
- Can learners track and learn *both* structures?

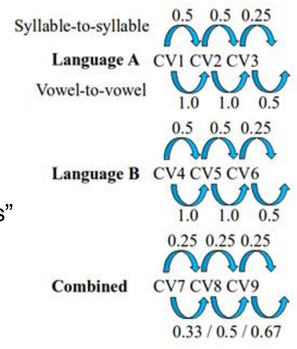
In the present study, we parameterize:
- Number of transitions between languages
- Duration of each segment of language.

Languages

Two languages (A and B) defined by the transitional probabilities (TP) between their syllables.

- Each language composed of two vowel-frames and six consonants (2 possible consonants per slot)
- Sixteen possible trisyllabic "words" in each language
- 50% overlap in syllable inventory between languages
- TPs distinguish within- and between-word units at vowel-frame and syllable level.

Combination of TPs between languages results in an unparseable statistical structure



Transitional probabilities defining the structure of each language. When combined, the TPs of each language result in a flat (uninformative) structure.

Continuous streams of computer-generated speech were generated by concatenating the Languages used by Gebhart and colleagues (2009), controlling for co-articulation.

Procedure

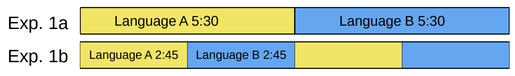
Learners were Penn State undergrads, English monolinguals

1. Exposed to a continuous speech stream of "foreign language"
2. Tested for word vs. part-word familiarity (32 item 2-AFC)
3. Language A vs. Language B performance are compared

Learners are exposed each language sequentially

You will listen to a recording of speech which sounds like a foreign language. There will be a quiz about what you've heard at the end…

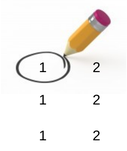
"pa ku be do ki tae po gi bae pa gu te da gu te..."



Learners are tested using word v. part-word judgements

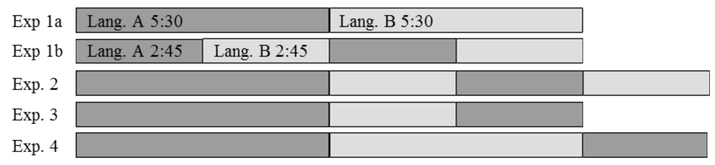
You will hear several pairs of speech segments. For each pair, circle 1 or 2 to indicate the one which sounds most familiar.

"da ku be… bae pa gu..."

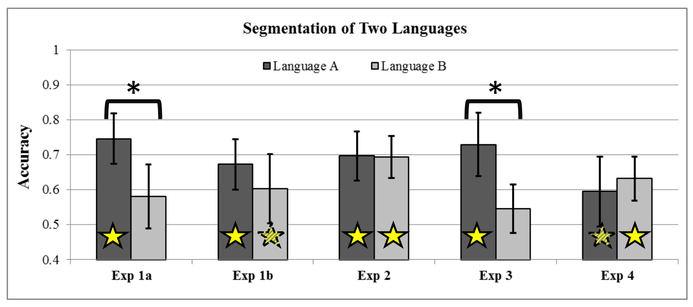


Four experiments were conducted in total
- Number of switches between languages and overall exposure to each language were varied between experiments.
- All experiments conducted with identical procedure, varying only the stream to which participants were exposed.

Results



Durations of each stream in four Experiments



Accuracy of Participant Judgments in Languages A and B in each Experiment

Experiment	1a	1b	2	3	4
N	17 (11F)	17 (12F)	20 (12F)	15 (13F)	17 (13F)
Age	19.6y	19.6y	19.9y	19.3y	19.3y

Experiment 1

Exp 1a: Primacy effect is **replicated** (see Gebhart et al, 2009)
Exp 1b: Primacy effect is **removed** (see Weiss et al, 2009).

Is Language A advantaged by an early entrenchment phase?
- *Entrenchment Account:* After 5:30 exposure, Language A has been learned and this representation is resistant to modification unless counter-evidence is overwhelming
- *Switching Account:* One switch from Language A to Language B is an insufficient cue to a context change, resulting in failure to learn B

Experiment 2

- Languages A & B significantly exceed chance
- Languages do not significantly differ from each other

No Primacy effect observed → Entrenchment is insufficient when additional switches occur (even with duration advantage for A)

Experiment 3

- Language A performance significantly greater than Language B (p=0.02, Primacy effect returns)
- Language B is not significantly greater than chance.

Decreasing the number of switches and duration of Language B restores the Primacy effect. Are two switches sufficient to eliminate the Primacy effect?

Experiment 4

- Performance in Language A significantly decreased from Exp 1
- Languages do not significantly differ from each other in learners' performance: Primacy effect again removed

A second switch is sufficient to eliminate Primacy, but neither language is learned. This interference effect resembles Weiss et al's (2009) results wherein neither language exceeds chance.

Summary

- Primacy effect achieved only in Experiments 1 & 3
  - When Language A has equal or greater exposure
  - AND two or fewer switches
- Language B learning was often marginally significant (see Exp 1 & 3), suggesting that even in Primacy conditions, some learning of Language B may have occurred.
- Experiment 4 increased A duration with only two switches, resulted in some learning of B (with reduced A learning)
  - Does this result suggest partial learning of each language and/or confusion between the languages?
- **Experiment 2 shows that when cues (switches) are provided, *both* languages can be learned at Primacy (Exp 1a) level, with only additional Language A duration.**

Discussion

- *Transitions* between languages (or switches) seem to modulate the *detection* of the second pattern (Exp 1b, 2, 5)
- This detection problem resembles a Bayesian approach (see Qian et al, 2012):
  - Learners must contend with variation in *any* signal
  - Prior probabilities for 1, 2, or more structures adjusted based on events that highlight variation (e.g., switches)
  - Structures are inferred based on the chosen model
    - 1 or 2 contexts
    - 1 context: ignoring 2nd language as natural variance
- Statistical learning is relies on inference about structures
  - Learners can use implicit & explicit cues
  - Statistical learning is *more* than co-occurrence
- Problems in early bilingual acquisition are tractable in statistical learning by inference about variability and underlying structures

Contact:
Benjamin Zinszer – bdz107@psu.edu / Dan Weiss – djw21@psu.edu

The Comparative Communication Laboratory @ Penn State | Center for Language Science