Learning phonological regularities across modalities*

Anne-Michelle Tessier and Marnie Krauss, University of Alberta
{amtessier, mkrauss}@ualberta.ca

Overview

*Main question: How well do adults learn novel phonotactics from brief **perceptual** (auditory) exposure? Does this exposure affect later **production** tasks like speeded repetition?*

Relatedly: *Does the artificial learning in this task resemble natural language learning in any way?*

The previous study: Onishi , Chambers and Fisher (2002)

The present studies: extending the methodology
Results: the influences of training and wordhood
a lack of natural class or generalization effects

Speculation: how artificial and natural phonological learning may differ: ability to generalize to new modalities (perception vs. production)?

1. About the nature of artificial language learning

The methodology: training participants (usually adults) on a small set of invented linguistic data, and then testing their resulting knowledge:

... what and how much they have learned...
... which patterns they have learned better than others...

*Special thanks to Pat Bolger for technical help and discussion, as well as to Andries Coetzee, Sara Finley, Shigeto Kawahara, Robert Kirchner, Gary Libben, Chelsi Mitchell, Kris Onishi, Anamaria Popescu, Laura Teddiman and the Centre for Comparative Psycholinguistics at the University of Alberta. Thanks also to Kyle Chambers, Lyn Frazier, John Kingston and Joe Pater for help in a previous incarnation of this research. Despite all this help, any errors in this talk are ours alone.

The rationale: (i) ability to control irrelevant factors (ii) experimental ease

A sample of recent linguistic research using artificial language learning: Saffran et al (1996); Pater and Tessier (2003); Pycha et al (2003); Wilson (2003), (2006); Carpenter (2005); Peperkamp and Dupoux (2006); Morrison (2006).

A question that is asked about artificial language learning:

"Is this learning *grammatical* learning? How do we know?"

Issue 1: People are very good at noticing and remembering a lot of *token-specific* knowledge, at some level (see, e.g. Chambers, Fisher and Church 1999, Church and Schacter 1994, Goldinger 1996, Nygaard and Pisoni 1998, Palmieri et al 2003, and many others).

Issue 2: What is the role of L1 knowledge in artificial language learning? Can we ever wipe the slate clean in a lab context?

The focus of this research program is **generalization**, in particular here:

• Generalizing from one modality to another
• Generalizing from one segment or natural class to another

What we will suggest today:

• Generalizing across modalities is, all things considered, very hard
• Getting away from L1 knowledge (or lexicons) is also hard

• Speculatively: perhaps the fact of how hard this generalization is may provide interesting fodder for the connections between artificial and natural learning
• More concrete speculation: perhaps a major difference between natural and artificial learning is the ability to *generalize across modalities*

2. Onishi et al (2002)

2.1 Procedure

Training: Listen to 120-130 CVC syllables over headphones, and rate each for clarity (24-26 words, 4-6 tokens of each)

Distractor: 28 two-digit addition problems

Testing: Listen to 98 CVC syllables over headphones, and repeat each one "as quickly as possible without making errors" into a microphone with a voice-activated response key

2.2 Materials

Experiment 1:
Onset and coda segments chosen from two groups that "could not be differentiated by a single phonetic feature or set of features"

1) Group 1: [b, k, m, t]
Group 2: [p, g, n, tʃ]

In addition:
onset [f] and coda [s] tokens in both lists
vowels: [æ] and [ɪ]

2) Pattern One[1] / Pattern Two

	Pattern One		Pattern Two
1.	bɪp	3.	pɪb
	mæn		næm
	tætʃ		tʃæt
2.	bæg	4.	gæb
	kæs		pæs
	bæp		pæb
	mɪn		nɪm
	tɪtʃ		tʃɪt
	bɪg		gɪb
	fɪg		fɪt

[1] The s-final and f-initial tokens are not necessarily the ones used in the experiment – the description of these materials in the text does not say which other consonants were combined with these segments, although they surely matched the rest of the lists's restrictions

Each participant was trained on one list, and tested on all four.

Lexical factor:

English words made up 49% of the Exp1 materials and 47% of the Exp3 materials. Including wordhood (real vs. nonce) as a factor in the model of the results 'revealed no main effect or interactions – all F < 1.46, p > 0.24.)

Author's conclusion: "Apparently, at least in this task in which words and non-words were freely mixed, lexical status had no effect on the phonotactic learning that we measured".

2.3 Results

- Reaction time (RTs) measured from stimulus offset to response onset.
- Mean RT scores reported exclude responses below 250 or above 1500 ms, and all those more than 2 SD beyond the subject's grand mean (mean = 4.2 per subject)

3) Testing, for one group in Onishi et al's Experiment 1

	List 1.	List 2.	List 3.	List 4.
familiar	bɪp	bæp	pɪb	pæb
licit	bæp	bɪp	pæb	pɪb
illicit	pɪb, pæb	pɪb, pæb	bɪp, bæp	bɪp, bæp

4) Mean reaction times (SD) for each condition

	familiar	licit	illicit
Exp 1	336 (131)	341 (132)	**350**(128)
Exp 3	**333** (117)	343 (122)	347 (119)

• In Exp 1, novel but licit syllables were repeated faster than illicit ones ($t(39) = 2.03$, $p < 0.05$)
• No difference in speed between familiar and novel licit ones ($t(39) = 1.28$, $p = 0.21$)

So: when subjects correctly repeated a word, they were significantly faster when it conformed to the onset/coda phonotactics of training

Overall:
• subjects can notice phonotactic patterns with limited exposure that affect linguistic performance in another modality.

• ... although the effect is pretty small (10 milliseconds!)

• real vs. nonce wordhood does not appear to affect speeded repetition in this paradigm

3. The present studies: Experiment 1

3.1 Materials

5) Group 1: **stops** [b, p, d, t] plus [n]
Group 2: **fricatives** [v, f, z, s] plus [m]

3.2 Procedure (modeled on Onishi et al 2002)
Experiment was run using DMDX.

Participants:
40 undergraduate students at the University of Alberta, all native English speakers with no reported hearing problems. Paid for participation.

Training:
• Listened over headphones to 24 CVC syllables, each 5 times for a total of 120 items.

• Rated each token on a three-way scale of clarity: 1 'unclear', 3 'average in clarity', 5 'very clear'.

• Training took approximately 15 minutes.

Distractor: 28 two-digit addition problems

Testing:
• Listened to 128 CVC syllables over headphones (including the familiar 24) and asked to repeat each one as quickly as possible without making errors into a head-mounted microphone

• DMDX recorded each response (.wav file), and RT between *offset* of stimulus and onset of response

3.3 Materials

6)

	SVF Pattern		FVS Pattern
1.	bæs	3.	sæb
	tæz		zæt
	dɪv		vɪd
	pɪf		fɪp
	næv		væn
2.	dæs	4.	sæf
	pæz		zæp
	bɪf		fɪb
	pɪv		vɪp
	tɪm		mɪp

Lexical factor:

Not controlled for (given Onishi et al's findings)
SVF: 21% real English words
FVS: 37% real English words

3.4 Predictions

- licit < illicit RTs, reflecting some internalization of the test materials
- no expected effect in error rates
- no expected effect of wordhood

4. Exp 1 Results

First 8 subjects had to be eliminated due to technical problems: 8 subjects per list remain, for a total of 32.

Our measure of interest: **RT**s, calculated only across all correct responses (excluding both incorrect repetitions and time-outs)

- RT trimming based on Onishi *et al*: excluded responses above 1500ms, below 100ms[2], and more than 2 SDs away from subject's grand mean

7) Word types for Experiment 1 Testing

Condition	Familiar	Licit	Illicit
List 1	bæs	pɪf	sæb, fɪp
List 2	pɪf	bæs	
List 3	sæb	fɪp	bæs, pɪf
List 4	fɪp	sæb	

Post-experiment debriefing
- absolutely no conscious recognition of the training data's pattern

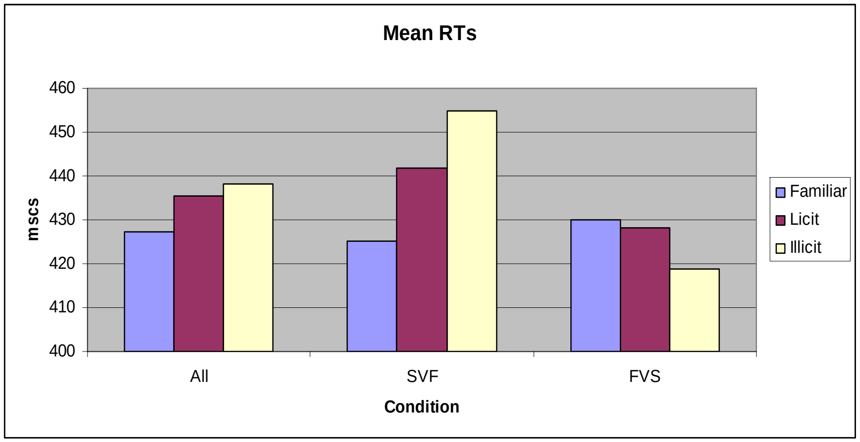
[2] Onishi et al's low cut-off was 250ms, but their RTs were recorded from the *onset* of target stimuli, not the *offset* as in the present study.

4.1. Raw data:

8) Mean RTs (SDs)

	Familiar	Licit	Illicit
All	427.3(117.4)	435.5(123.1)	438.1(132.7)
SVF	425.1(117.7)	441.9(119.8)	454.9 (114.0)
FVS	429.9(121.5)	428.1(130.9)	418.9(153.5)

9) Mean RTs for all subjects



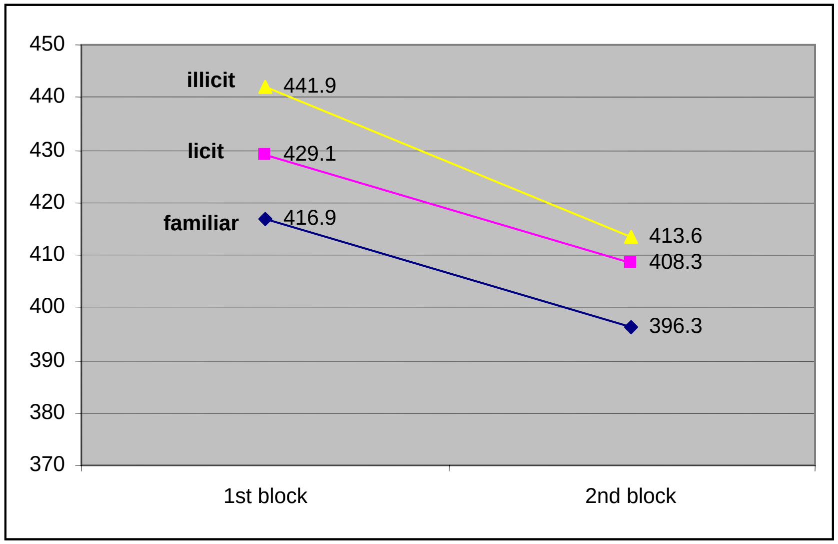
Only significant result: SVF familiar < licit (t(*15*) = 3.01, p < 0.01)
Trend: SVF licit < illicit (t(*15*) = 1.52, p = 0.075)

Overall: Unexpectedly, no clear evidence of training

4.2 Learning throughout testing?

As suggested by Onishi et al: breaking down results into two blocks:[3]

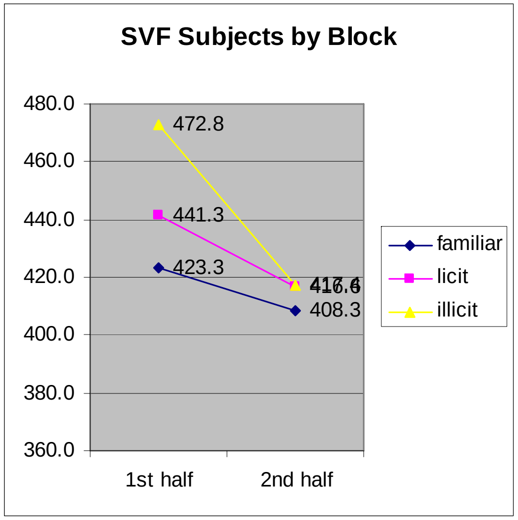
10) Mean RTs for all subjects, by block



Looking at all subjects: responses on each word type got faster over time ($p =$ or < 0.05), but no interaction between word type and block

Comparing the two blocks for just SVF subjects:

11)a)



For illicit words:

SVF 1st block > 2nd block ($t(15) = 2.25$, $p < 0.02$)

Contrast analysis of Illicit blocks 1 and 2 in both SVF and FVS shows a significant interaction of training condition and wordtype ($t(23) = 1.94$, $p = 0.032$)

In comparison: FVS subjects showed a trend for speeding up on familiar and licit words between first and second blocks (p values approx. .08)...

- But they remain at a consistent speed on illicit words (1st block: 410.8ms; 2nd block: 409.9 ms)

Overall: SVF subjects seem to have started treating illicit words as licit, as expected, but FVS subjects clearly did not.

3 Four subjects' data was not included in this comparison, because they made more than 25% errors on in one or more condition in one block.

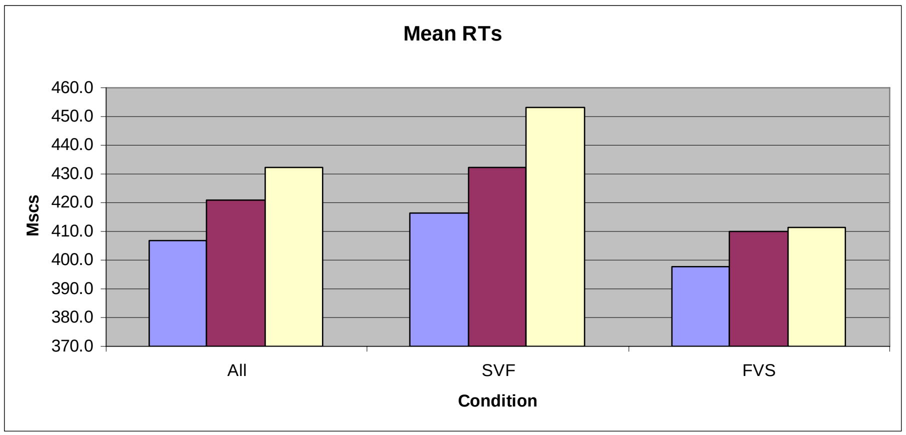
4.3 Testing for a lexical effect?

Recall: SVF: 21% real English words
FVS: 37% real English words

12) Mean RTs just for **nonce** words:

	familiar	licit	illicit
All	407 (125.7)	421.0 (137.4)	432.3 (147.7)
SVF	416.2(127.5)	432.2 (131.9)	453 (135.3)
FVS	397.9 (127.4)	409.8 (146)	411.5 (160.9)

13) Mean RTs *for nonce words only*
Nonce words only: familiar-licit-illicit wordtypes



SVF training: • familiar < licit (t(15) = 2.23, p = 0.02)
• AND licit < illicit (t(15) = 2.62, p < 0.01)
FVS training: • marginal familiar < licit (t(15) = 1.56, p = 0.069)
• but still licit = illicit

Overall In nonce words only, SVF training did produce the expected effect of training – but not so for FVS training.

4.4 The overall SVF < FVS effect (a confound)

Looking at the RTs without the bias of training-based labels (e.g. familiar)

14) Grand mean, across all subjects and items

SVF words	FVS words
426.7 (133.6)	**442.8** (117.6)

Ignoring everything else: SVF words were repeated faster than FVS words (t(29) = 2.17, p = 0.019).

• No doubt this is part of the explanation for the weird FVS results

•... But it cannot be the whole story (e.g., it cannot explain the difference in learning over blocks between SVF and FVS training conditions.)

Summary of Experiment 1

1. In Experiment 1: with real and nonce words combined, RTs did not show an advantage of licit over illicit words in speeded repetition
2. Looking at changes over the course of testing and at nonce words only, SVF conditions show evidence of learning as in Onishi *et al*
3. Overall SVF < FVS items, confounding results
4. Lexical status (real vs. nonce words) can indeed interact with the influence of training

5. Experiment 2

5.1. Design, Materials, Procedure

Methodology and participant population was identical to Exp 1, materials differed in several ways

a) Different natural classes

• defined by place rather than manner: labials vs. coronals (to avoid microphone-triggering confound)

- In addition: velar stops appeared in both onset and coda positions
- To be able to build enough items, used 8 different vowels

15) Group 1: **labials** [p, m, f, v] plus [k, g]
Group 2: **coronals** [t, n. s, z] plus [k, g]

16) **Training:**

Lab-V-Cor

pɛs	gɪz
miz	mok
vɪt	ket
fun	vaeg

Cor-V-Lab

tʌp	kof
nɪm	sɛk
sif	zag
zaev	guv

(Note: 4 lists were constructed, two for each training condition, but we will report on only one of each groups tested here.)

b) Two different voices

• To make the clarity rating task more reasonable

c) New segments in testing

• in testing, participants heard items with a novel labial [b] and novel coronal [d] in coda position, and a velar in onset[4]

• These were designed to see whether novel members of a restricted training set (labials or coronals) would pattern with training segments

[4] The reverse set of items, with novel segments in the onset position, were not used because too few non-words could be constructed.

17) **Testing: generalization items**

kɪb	kob
gab	gib
ked	kud
gud	gaed

d) Only nonce words [5]

• given findings of Exp 1

18) Word types in Exp 2 Testing

Condition	**Familiar**	**Licit**	**Illicit**	**Gen-Licit**	**Gen-Illicit**
Lab-V-Cor	pes	faez	tʌp, guv	kud, gaed	gib, kob
Cor-V-Lab	tʌp	zim	pɛs, faez	gib, kob	kud, gaed

Procedure *(roughly the same as Exp 1)*

Training:
• 32 CVC syllables, including 16 with velars
• 4 blocks of repetition: 2 times with each of 2 female voices; 128 tokens in total

Testing:
• 150 CVC syllables (including the familiar 32 and the 22 generalization items) and asked to repeat each one as quickly as possible without making errors

5.1.1.Predictions

• licit < illicit RTs, reflecting some internalization of the test materials

• if new segments treated as members of training set:
generalization-licit < generalization-illicit

[5] As much as possible. ... Some syllables used were borderline – nicknames, slang, product names (e.g. ‘bick’) or very low frequency/archaic terms.

5.2 Results

- 24 subjects run to date
- two subjects excluded, due to high error rates or equipment error
- reported here: 9 subjects in Lab-V-Cor; 10 in Cor-V-Lab

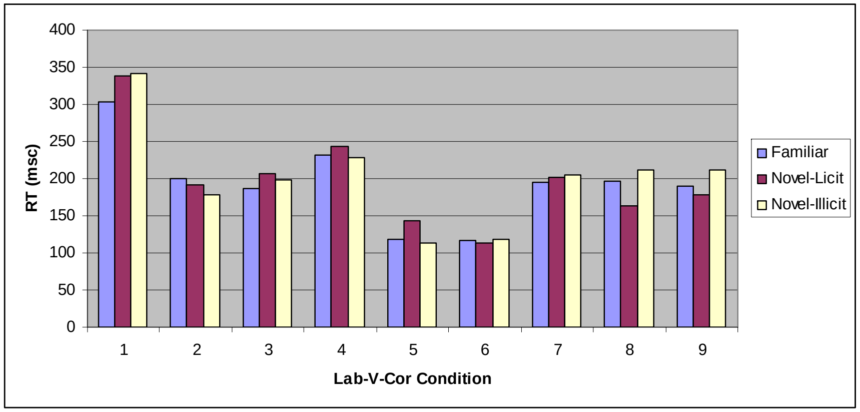
5.2.1 Basic items in testing: familiar, novel-licit and novel-illicit

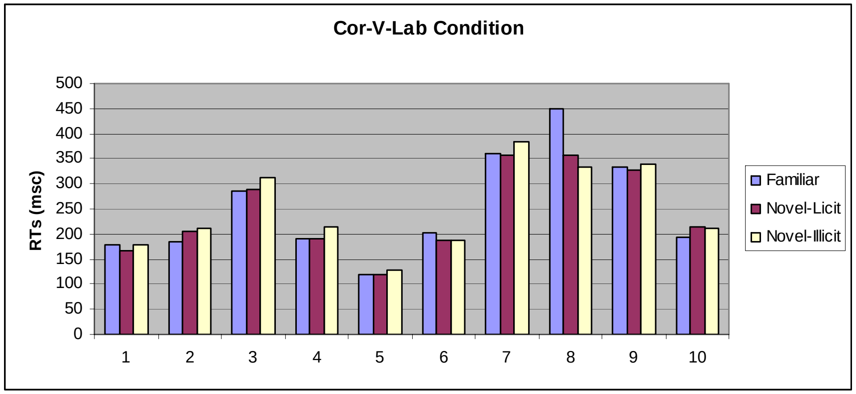
19) Mean RTs (SDs)

	Familiar	Licit	Illicit
All	222.9 (86.4)	220.5 (77.1)	226.6 (78.9)
Lab-V-Cor	193.1 (55.9)	197.7 (64.6)	200.7 (66.6)
Cor-V-Lab	249.7 (102.3)	241.0 (84.8)	250 (84.9)

Overall: no significant patterns to the results

20)a)





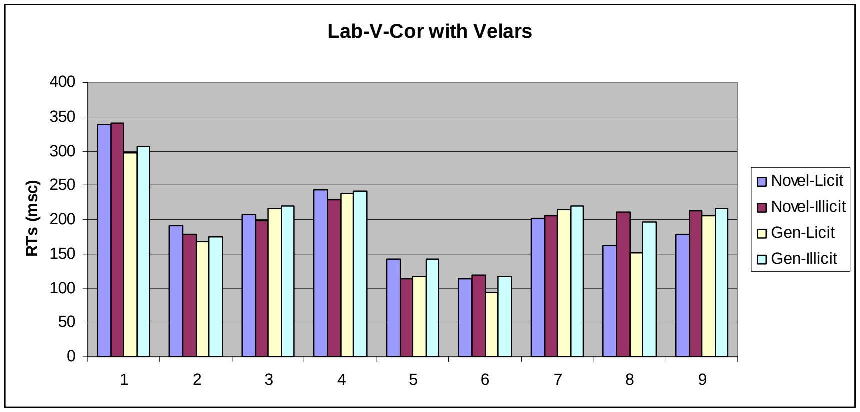
5.2.2 Generalization items in testing: gen-licit vs. gen-illicit

21) Mean RTs (SDs)

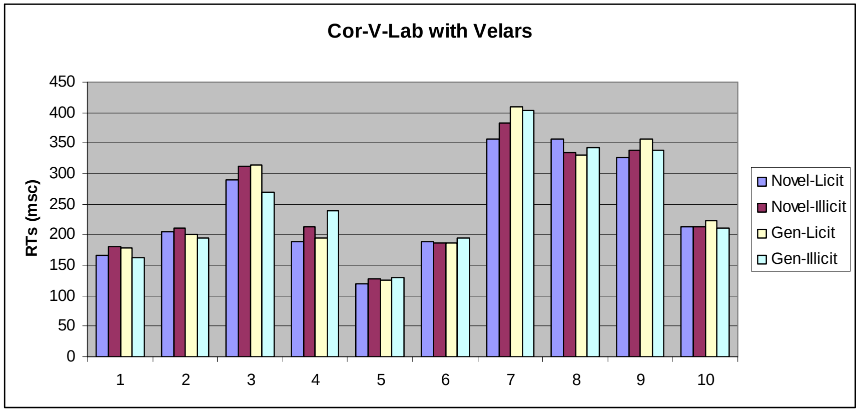
	Gen-Licit	Gen-Illicit
All	222.2 (84.5)	227.1 (76.1)
Lab-V-Cor	189.0 (63.0)	203.9 (55.5)
Cor-V-Lab	252 (93)	248.1 (88.5)

Overall: The only (unexplained) pattern is that Lab-V-Cor participants repeated gen-licit items faster than basic novel-but-licit!

22)a)



22)b)



Summary of Experimental Results

After brief listening exposure to stimuli with positional restrictions:

- some participants might repeat novel syllables slightly faster,
- ... with only some particular restrictions
- ... with only nonce words
- ... with no advantage for restrictions using natural classes
- ... and no generalization to similar segments

In other words: the effect of brief listening on speeded repetition is ephemeral at best

6. Speculation about experimental tasks and learning

Observation: participants can succeed quickly at other tasks in which novel phonological restrictions are learned and tested:
using *either production or perception* – only one or the other

Slips of the tongue paradigms: all production See esp. Dell et al (2000), Dell and Warker (2004), Goldrick (2004); Taylor (2003), Taylor and Houghton (2005); Warker and Dell (2006).

Method: Participants repeat quadruples of nonsense syllables with novel phonotactic restrictions, first slowly and then to a faster metronome beat

Result: slips of the tongue are mediated by production experience – participants' slips of the tongue show the influence of an experimentally-induced version of the 'syllable position' effect

First order restrictions, like 'f only appears in onset' (exactly like the ones in our own experiment) begin to affect slips very quickly – apparently, as fast as in 9 trials (Taylor 2003, Dell and Warker 2006)

Listening to foreign-accented speech: all perception (auditory/ visual processing) See e.g. Clarke and Garrett (2004), Clarke and Luce (2005), Norris, McQueen and Cutler (2003)

Method in Clarke and Garrett: participants listened to sentences in heavily-accented speech (e.g. Spanish or Mandarin-accented English), and after each sentence saw a visual probe word and were asked to judge if it matched the last word of the target sentence.

Result: improved speed of processing occurs very quickly – over only sixteen trials.

Children learning novel phonotactics: all perception See review of results in Fisher, Church and Chambers (2004)

With materials like the Onishi *et al* stimuli and our own, but with age-appropriate methods (head-turn-style testing): ***infants*** do well.

In contrast:

In the present study using natural classes did not help, and may even have hindered participants

... At any rate, evidence of learning was very weak (mildly present in Exp 1, and so far completely lacking in Exp 2)

The speculative synthesis of these results

***Hypothesis**:* One crucial way that adult artificial learning differs crucially from natural learning is that it does not work in both modalities at once.

From brief exposure to novel phonologies:

- we can learn to produce OR perceive
- but it is difficult, bordering on the impossible, to integrate from one modality to the other

Recall however that Onishi *et al* did also get an effect of learning

- as they discuss, this kind of 'transfer' makes up a plausible and probably necessary part of natural L1 phonological learning (see e.g. Pater 2004, Boersma 2006)

Understanding the kinds of limitations and possibilities for novel learning across modalities, and what they tell us about natural vs. artificial language learning, are still very much issues for future research.

Thank you very much.

7. References

Carpenter, Angela (2005). BUCLD. "Acquisition of fa Natural vs. an Unnatural Stress System." In A. Burgos, M. R. Clark-Cotton and S. Ha (eds.), *Proceedings of BUCLD29.* Somerville, MA: Cascadilla Press. pp. 134-143.

Chambers, K. and K. Onishi (2004) "A vowel is a vowel is a vowel" *Proceedings of BUCLD29.*

Clarke, C. and M. Garrett (2004). 'Rapid adaptation of foreign-accented English.' *JASA* 116 (6).

Clarke, C. and P. Luce (2005). 'Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization.' In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception* (pp. 23-26), London.

Dell, G., K. Reed, A. Adams and D. Meyer, (2000). "Speech Errors, Phonotactic Constraints, and Implicit Learning: A study of the Role of Experience in Language Production." *Journal of Experimental Psychology: Language, Memory and Cognition,* 26(6): 1355-1367.

Dell, G. and J. Warker (2004) "The tongue slips into (recently learned) patterns." To appear in a festchrift for S. Nootebaum.

C., Church, B., & Chambers, K. E. (2004). Learning to identify spoken words. In D. G. Hall and S. R. Waxman (Eds.) Weaving a lexicon (pp. 3-40). Cambridge, MA: MIT Press.

Fisher, C., B. Church and K. Chambers (2004). 'Learning to identify spoken words'. In D.G. Hall and S.R. Waxman (eds.) *Weaving a lexicon.* Cambridge, MA: MIT Press.

Goldrick, M. (2004) "Phonological features and phonotactic constraints in speech production." *Journal of Memory and Language* 51: 586-603.

Morrison, Geoffrey S. (2005). "Phonetic naturalness and phonological learnability." Paper presented at the 13th Manchester Phonology Meeting (mfm13).

Norris, D. J. McQueen and A. Cutler (2003). 'Perceptual learning in speech'. Journal of Cognitive Psychology v. 47, 204-238.

Onishi, K., K. Chambers and C. Fisher (2002). "Learning phonotactic constraints from brief auditory experience *Cognition* 83: B13-N23.\

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). ''Episodic encoding of voice attributes and recognition memory for spoken words,'' *Journal of Experimental Psychology: Learning, Memory, and Cognition,* v. 19, 309–328.

Pater, Joe and Anne-Michelle Tessier (2003). "Phonotactic Knowledge and the Acquisition of Alternations". In In M.J. Solé, D. Recasens, and J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona. pp. 1777-1180.

Peperkamp, Sharon and Emmanuel Dupoux (2006). 'The Role of Phonetic Naturalness in Phonological Rule Acquisition.' In D. Bamman, T. Magnitskaia and C. Zaller (eds.) *Proceedings of BUCLD30.* Somerville. MA: Cascadilla Press. pp. 464-475.

Pycha, Anne, Pawel Nowak, Eurie Shin and Ryan Shosted (2003). "Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony." In G. Garding and M. Tsujimura (eds.) *Proceedings of WCCFL22.* Somerville, MA: Cascadilla Press. pp. 423-435.

Saffran, Jenny, Elissa Newport and Richard Aslin (1996). "Word Segmentation: The Role of Dsitributional Cues". *Journal of Memory and Language* 35: 601-621

Taylor, C. F. (2003). *The acquisition of phonemic constraints: implications for models of phonological encoding.* Unpublished doctoral dissertation, University of Wales, Bangor.

Taylor, C. F. & Houghton, G. (2005). Learning artificial phonotactic constraints—time course, durability, and relationship to natural constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31*,* 1398–1416.

Warker, J. and G. Dell (2006). 'Speech errors reflect newly learned phonotactic constraints'. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 32: 387-398.

Wilson, Colin (2003). "Experimental Investigation of Phonological Naturalness". In G. Garding and M. Tsujimura (eds.) *Proceedings of WCCFL22.* Somerville, MA: Cascadilla Press. pp. 533-546.

Wilson, Colin (2006). "Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization." To appear in *Cognitive Science.*