Statistics All the Way Down: How is Statistical Learning Accomplished Using Varying Productions of Novel, Complex Sound Categories?

Lauren L. Emberson (lle7@cornell.edu)
Ran Liu (ral2012@med.cornell.edu)
Jason D. Zevin (jdz2001@med.cornell.edu)
Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University
Box 140, 1300 York Ave, New York, NY, USA, 10065

Abstract

For statistical learning to aid in language learning, learners must resolve statistical information along multiple dimensions of the same linguistic signal. Given that infants show evidence of lexical knowledge while they are still learning how to categorize speech, infant learners are likely presented with at least two statistical learning problems simultaneously. In an effort to approximate this scenario, we presented adult participants with multiple exemplars of sounds from 4 experimenter-defined categories. These sounds were novel and thus, adult have not developed specialized processing for these sounds. Stimuli were presented in a regular, continuous stream containing statistical structure between sound-category types with variable exemplars (i.e. pairs of sound categories but with variable exemplars of each category presented instead of just one). Participants were tested for familiarity with high probability pairs. We found that participants can learn from statistical structure based on varying exemplars of novel sounds but they learn based on the perceptual grouping biases that they bring into the experiment and not based on the experimenter-defined categories (groupings they would have to form *ad hoc* in the experiment). We discuss these results in relation to language learning.

Keywords: Statistical learning; unsupervised learning; language learning and development; speech categorization; cognitive development; auditory processing.

Introduction

Throughout development, humans are highly sensitive to statistical regularities in the environment. From these regularities, it is possible to learn a large amount about the structure of the world without explicit feedback or innate knowledge. Statistical learning has been intensely studied in relation to one of the most formidable tasks that humans face: learning language. It has been established that statistical information can aid infants in many aspects of language development including speech categorization, lexical development, and syntactic processing, even in the first year of life (see Thiessen & Saffran, 2007 for a review).

For a colloquial example, take the phrase "pretty baby" (Saffran, Newport, & Aslin, 1996) which would typically be produced as a continuous utterance /prI'tibeI'bi/. In the ambient language, the transitional probabilities (as well as co-occurrence frequency) are higher within syllables of words than between the syllables at the boundaries of words. Not only have infants have been shown to be sensitive to these transitional probabilities as early as 2-months of age (Kirkham, Slemmer, & Johnson, 2002), syllables linked by high transitional probability are more likely to be used as lexical labels (Graf-Estes, Evans, Alibali, & Saffran, 2007). Thus, it is thought that this learning ability is likely to contribute to lexical development, characterized in part by the word explosion beginning around 14-months (Thiessen & Saffran, 2007).

However, in the infant's acoustic environment, these transitional probabilities are necessarily accumulated over many, many instances of hearing different productions of the same continuous utterance (e.g. "pretty baby" or /prI'tibeI'bi/). Across these utterances, there is a large amount of acoustic variability that functionally belong to the same speech category.[1] This variation is thought to be dealt with through the process of speech categorization. However, even though infants in the first year *begin* to preferentially discriminate the acoustic contrasts employed in their ambient language (Werker & Tees, 1984), the solidification of speech categories continues well beyond infancy (e.g. Hazan & Barrett, 2000).

Thus, the process of speech categorization has a largely overlapping developmental time-course to statistical learning of transitional probabilities and the early stages of language learning. In order for infants to learn words based on the statistical information in their ambient language, they are likely presented with at least two statistical problems simultaneously: infants have to learn that syllables cohere to form a word while simultaneously resolving that the many variable productions are functionally equivalent in their native language.

In previous studies examining statistical learning of transitional probabilities, participants are exposed to a corpus that consists of acoustically identical repetitions of sounds. This is markedly different from the multiple varying productions as exist in infants' early language experience. In these experiments, stable physical properties across multiple presentations eliminates the problem of categorization across multiple, varying productions of speech. Likewise in adulthood, speech processing is largely robust to variations across speech productions.

[1] There are many other sources of information that vary across productions, including contextual information, visual environment, and interaction with the caregiver, all of which have been shown to modify cognitive processing in infancy and later in development and thus will alter the informational content of each utterance.

In an attempt to approximate the task facing an infant attempting to form lexical knowledge while lacking completed speech categorization, the current experiments investigate statistical learning using multiple, varying exemplars of novel, complex sound categories in adults. The stimuli were adapted from the training study of Wade and Holt (2005) which employed a video game paradigm to implicitly train listeners to learn categories of novel, spectrotemporally complex non-speech stimuli. These sounds are carefully designed to capture some of the spectral complexities that exist within natural speech categories without sounding speech-like. For current purposes, it is essential to note that adult participants have not heard these sounds before nor have they undergone any experiences that would result in the formation of functional categorization of these sounds.

In the current experiments, we employed the 4 experimenter-defined sound categories from Wade and Holt (2005) and grouped these sound categories into pairs (i.e. pairs of sound categories were linked by high transitional probability). We presented 6 different exemplars of these novel sounds. Thus, participants have to group sounds across multiple productions in order for the transitional probabilities to be reliable. This is similar to what an infant faces when learning transitional probabilities without fully developed functional sound categories.

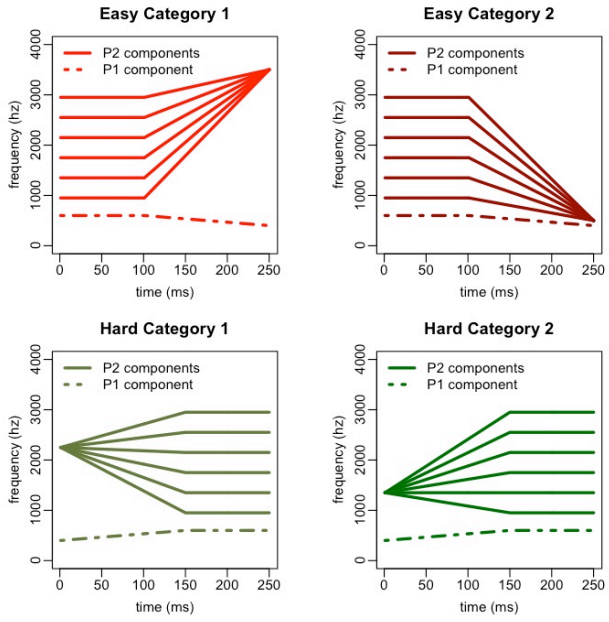


Figure 1: Schematic diagram of the spectrotemporal properties of the stimuli employed in all experiments. Each sound has two components: P1 (constant over all stimuli in a category) and P2 (varies for each stimulus).

Experiment 0: Naïve Perceptual Biases

Adapted from Wade and Holt (2005), six stimuli from each of four experimenter-defined categories were used. All sounds were designed to have two spectral peaks, P1 and P2 with both steady-state and gradually changing portions, similar to syllables containing a vowel and a semivowel or liquid, (for a schematic diagram of the four categories of stimuli, see Figure 1).

The two "Easy" categories comprise stimuli that begin with a steady state period and then either rise or fall. These were designed to be easy to learn because they are reliably discriminable by the direction of the transition period. The two "Hard" categories begin with a transition, followed by a steady state. Note that stimuli from both Hard categories should, therefore, be highly discriminable from the Easy stimuli. The contrastive cue that distinguishes Hard categories is the onset frequency of the transition period. Both contain rising and falling frequency patterns and completely overlap in their steady-state frequencies, and thus, only a higher-order interaction between these two cues creates a perceptual space in which the two hard categories are discriminable (see Wade & Holt, 2005 for a comprehensive discussion and http://www.psy.cmu.edu/~lholt/php/gallery_irfbats.php to hear the sounds).

Because the stimuli and experimenter-defined categories are spectrotemporally complex and completely novel to participants, two consequences follow: 1) participants do not have specialized processing for these sounds like they do with speech processing. Therefore, we believe there is no *a priori* categorical perception of these stimuli (see Wade & Holt, 2005). 2) However, it is unlikely that experimenter-defined categories will be perceived to be equally distinct. To investigate the perceptual biases participants bring to the learning tasks, we asked naïve participants to perform a perceptual similarity judgment.

Methods

Participants 28 students participated in the current study. All participants reported in this paper were undergraduates at Cornell University who participated in exchange for course credit. Participants were asked to report any auditory, visual, or neurological deficits via post-experimental questionnaire; no participants reported any such deficits.

Stimulus presentation All sounds were presented using over-the-ear headphones (Sony MDR-V150) at a comfortable, above-threshold volume. Instructions and stimuli were presented using PsyScope X B53 on MacMini computers. During sound presentation, participants observed blank, white screens on 17in CRT monitors. All sounds were presented for 300ms. Each trial began and ended with 500ms of silence and the two sounds were presented separated by a pause of 500ms.

Similarity Judgment After hearing both sounds, participants were asked to report how similar the sounds were on a scale of 1 to 4 (1 = the same and 4 = completely different) on a keyboard. Participants were given an unlimited amount of time to make their responses.

For practical purposes it was necessary to limit the number of trials by partitioning the full set of 24 exemplars (6 from each of the 4 categories) into two subsets; one subset contained exemplars 1, 3, 5 from each category and the other subset contained exemplars 2, 4, 6. Half the participants performed similarity judgment on one subset, the other half on the other subset.

Results and Discussion

Similarity for each stimulus pair was computed for each participant. Similarity judgments for each contrast were analyzed directly in a one-way ANOVAs with subject (F_1) as a random factor ($F_1(1,27) = 1982.1$, $p < 0.0005$) revealing contrast (E1-E1, E1-E2, E1-H1, E1-H2, etc.) as a significant variable: $F_2(9,243) = 58.92$, $p < 0.0005$. Within category judgments were smallest, indicating the most similar judgments, for E1 and E2 judgments (1.45 and 1.53 respectively) and larger for H1 and H2 judgments (2.23 and 2.20 respectively) with a significant main effect of category in an ANOVA ($F(3,81) = 24.42$, $p < 0.0005$). These results reveal within category cohesion for both Easy categories indicating that they are "perceptually grouped" naïvely.

Both our graphical results (Figure 2) and the statistical analyses indicate that E1-E1 and E2-E2 exemplars are rated more similarly than any other exemplar comparison. However, participants do not rate H1-H1 or H2-H2 exemplar pairs more similarly than they do H1-H2 pairs suggesting that they do not treat H1 and H2 as two separate groups but as one single "perceptual group" separate from E1 and E2.

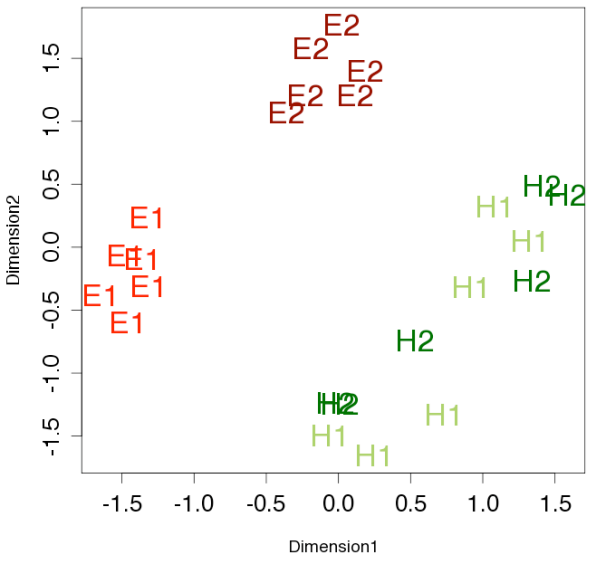


Figure 2: Perceptual Distance between all stimuli for all subjects. Similarity judgments were entered into MDS analysis with two dimensions.

Experiment 1: Statistical Learning Using Multiple Auditory Exemplars

After establishing the perceptual biases and groupings that participants bring to the corpus of sounds, we presented a new group of participants with the sounds in a continuous stream with statistical structure (i.e. pairs) defined over sound category (e.g. E1-H2). Critically, we presented one exemplar of an E1 sound of which there are 6 and one exemplar of an H2 sound, of which there are also 6. In other words, multiple exemplars from each sound category are presented.

In the current paradigm, statistical learning can take place based on specific exemplars (e.g. E1_1-H2_5 vs. E1_1-H2-2), consistent with previous findings, or over groups of multiple exemplars of sounds (e.g. E1-H2). We didn't anticipate that participants would learn based on the individual exemplars (of which there are 24), two specific exemplars (e.g. E1_1-H2-2) are only presented together twice during the entire familiarization stream. Thus, we focused our analyses on how learning takes place across groups of sounds.

Although the corpus we employed has four categories of sounds constructed *a priori* (see Introduction), results from Experiment 0 indicate that naïve participants do not equally distinguish all 4 categories of sounds. Thus, if participants are able to learn across multiple exemplars of sounds, it is possible that they could learn based on different perceptual groupings of stimuli. Specifically, we examined whether the pattern of behavioral results is consistent with participants learning based on different **levels of perceptual grouping**:

1) **Four groups of sounds (E1, E2, H1, H2)** based on the four experimenter-defined categories, but not distinguished by naïve participants;

2) **Three groups of sounds (E1, E2, H)** as seen in the naïve perceptual groupings. Thus exemplars of E1 and E2 sounds would be grouped separately and the two Hard categories (H1 and H2) would be treated as distinct from the two Easy categories but being indistinguishable from each other.

3) **Two groups only (E, H)** with the Easy categories (E1 and E2) being perceived as a single group and the Hard categories (H1 and H2) being perceived as a second group.

Based on how participants group the different exemplars (perceptual grouping) as well as the sound-pair assignment (discussed below), we made specific predictions as to whether or not participants would be able to demonstrate learning at test (see the table in Figure 4 for a summary of these predictions).

Methods

Participants 45 participants were recruited for this experiment. One participant was excluded for failing to complete the entire experiment.

Sound-Pair Assignment For each participant, the four categories are grouped into two pairs (e.g. E1-H2, H1-E2). We will refer to this as a *sound-pair assignment*.

Familiarization Each sound was presented for 300ms with a 115ms inter-stimulus interval (ISI) for a 415 stimulus onset asynchrony (SOA). All the exemplars from each category are paired with all other exemplars from the paired category of sounds twice resulting in a familiarization stream of 648 pairs of sounds constructed from 24 different exemplars from the four categories of sounds presented in randomized order (see Figure 3, top row).

Cover Task In order to encourage participants to pay attention to the familiarization stream without explicitly asking them to track the relationships between sounds heard, a cover task was employed which consisted of participants detecting 'soft' sounds by pressing the SPACE bar. A sound attenuated version of each exemplar was added into the familiarization six times resulting in 144 'soft' sounds out of the 1296 sounds presented. These were incorporated into the familiarization stream and thus did not disrupt the statistical structure of the stream. Participants were instructed that they would hear a stream of sounds and to press the SPACE bar when they heard the stream get quieter. They were also told that the task would last 7 minutes. Button presses within 1.6 seconds of presentation of the soft sound were considered a correct response.

Test for Statistical Learning After familiarization, participants were given a self-timed break and then told that they would be presented with two pairs of sounds separated by a long pause (1000ms) and after hearing both, they would be asked to report which pair of sounds is more familiar based on their previous task. They used the 'g' and 'h' keys to indicate which pair was more familiar. They were also told that no new sounds are being introduced and encouraged to go with their intuition or 'gut instinct'. The responses were self-timed.

Participants were given 48 test trials. In each trial, one pair was composed of two exemplars consistent with those in familiarization and the other was a *foil* that violated the statistical structure from familiarization. Foils were constructed based on the 4 experimenter-defined sound categories: if pair 1 is AB and pair 2 is CD then the foils are CA and DB. All exemplars were heard and each pair was paired with each foil an equal number of times and counterbalanced order.

Perceptual Similarity Judgment After completion of the SL test, participants were asked to perform a perceptual similarity judgment, as described in Experiment 0.

Results

Cover task results Participants responded correctly to the 'soft' sound with an average of 76% accuracy. We didn't exclude any participants based on Cover Task performance.

Perceptual Similarity Judgments: We did the same analyses on the perceptual similarity judgments as in Experiment 0 and found identical results indicating that participants have a stable perceptual bias in relation to the corpus of sounds throughout the experiment. An ANOVA with subject as a random factor (F_1) revealed a significant effect of category comparison (F_2 (9,387) = 65.13, $p < 0.0005$; $F_1(1,43) = 2501.1$, $p < 0.0005$) with the lowest average judgments for within E1 and within E2 trials (1.39 and 1.61 respectively). Within H1 and H2 trials were on average much greater (2.08 and 2.09 respectively). Within category comparisons also yielded a significant effect of category (F_2 (3,129) = 25.93, $p < 0.0005$).

Discrimination at test We first examined behavioral responses for evidence of learning for all participants together, regardless of which sound-pair assignment. When evaluated against chance performance (24 out of 48 or 50% performance), we find that overall, participants were able to reliably distinguish the category pairs heard during familiarization from foils: mean performance = 27.93, std = 6.65, $t(43) = 3.93$, $p < 0.001$.

Transitional Probabilities in Familiarization and at Test The experimental organization of the four categories of sounds creates an *a priori* set of transitional probabilities which are higher within pairs of sound categories (100% or 1.0 transitional probability within pairs) than between pairs.

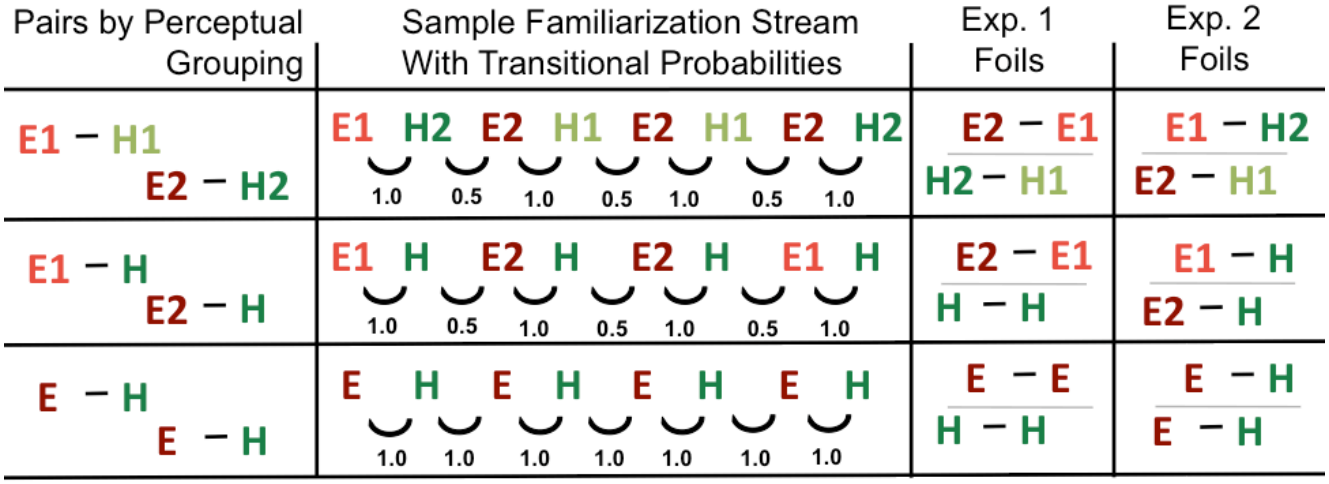


Figure 3: Differences in transitional probabilities and foils for both Experiment 1 and 2 across the 3 levels of perceptual grouping: E1, E2, H1, H2 (experimenter-defined categories); E1, E2, H (naïve perceptual groupings), and E vs. H (minimal naïve groupings).

However, because the statistics are defined at the level of *category* rather than *exemplar*, the statistical information could differ from the experimental design is participants' groupings differ from the *a priori* categories. Figure 3 illustrates how different perceptual groupings of multiple exemplars change transitional probabilities during familiarization and how different perceptual groupings will change transitional probabilities of the foils at test.

Using the transitional probabilities during familiarization and of the foils, we made predictions for the test conditions. Specifically, if the transitional probabilities between pairs are 1.0 during familiarization[2] and the transitional probability of the foils averaged less than the transitional probabilities between pairs during familiarization (i.e., less than 0.5), we predicted above-chance performance in discriminating learned pairs from foils. A summary of predictions is presented in Figure 4.

Analysis based on Sound-Pair Assignments Next, we divided participants into three subgroups based on sound-pair assignment. All possible sound-pair assignments were used in the experiments equally often. However, as seen in both the perceptual judgment of naïve participants and participants who have undergone familiarization, the perceptual space is not homogenously parsed for all categories. Thus, not all sound-pair assignments (which differ across participants) are equivalent. For the purposes of defining subgroups we equated both Easy and Hard categories and defined three subgroups as follows:

Sound-Pair Assignment 1 participants who had both Easy categories assigned to one pair and the two Hard categories assigned to the other (i.e. EE, HH).

Sound-Pair Assignment 2 participants who had Easy and Hard categories mixed between pairs but in consistent ordinal position in both pairs (i.e. EH, EH or HE, HE)

Sound-Pair Assignment 3 participants who had Easy and Hard categories mixed between pairs in different ordinal position (i.e. HE EH).

A one-way ANOVA revealed a significant effect of Sound-Pair Assignment ($F(2, 41) = 7.71$, $p = 0.001$). Performance for each group (see Figure 4): S-P Assign. 1 and S-P Assign. 2 reliably discriminated correct pairs from foils (S-P Assign. 1: mean = 28.5, std = 7.06, $t(13) = 2.37$, $p < 0.05$; S-P Assign. 2: mean = 31.8, std = 5.83, $t(14) = 5.18$, $p < 0.001$) whereas participants in S-P Assign. 3 failed to discriminate correct pairs from foils: mean = 23.53, std = 4.27, $t(14) = -0.423$, $p > 0.5$. Thus, we find differences in the ability of participants to distinguish foils from pairs depending on Sound-Pair Assignment. Comparing this pattern of results to our predictions (Figure 4), we find evidence that participants learned across multiple sound exemplars based on the perceptual groupings that they had prior to the experiment (results from Exp. 0).

[2] This is true in all cases except for perceptual grouping of less than 4 groups for the third Sound-Pair Assignment possibility.

Discussion

These results indicate statistical learning is possible across multiple exemplars of sounds from novel, complex categories. However, in order to accomplish this, participants relied on the perceptual groupings of these sounds that they bring into the task as demonstrated by the large effect of Sound-Pair Assignment. Our predictions, based on transitional probabilities during familiarization and at test, predict test performance across Sound-Pair Assignment only with the assumption that participants group the sound exemplars according to their initial perceptual biases and not according to either the experiment-defined categories or according to a two-way discrimination of Easy and Hard categories (perceptual grouping of E1, E2, and H; see Figures 3 and 4).

Experiment 2: Changing the Foils

We presented evidence in Experiment 1 that participants are able to learn statistical structure of novel, complex auditory categories and they learn based on their naïve perceptual groupings. In the current experiment, we change the foils that we employ at test to produce different predictions for learning across Sound-Pair Assignments (see Figure 4), thereby allowing us to further test our assumption that participant learn over multiple sound exemplars based on their naïve perceptual groupings of the stimuli.

Methods

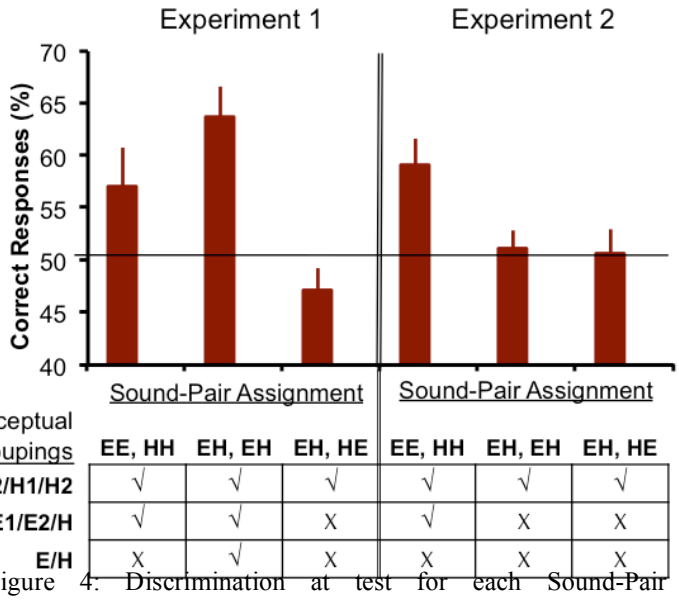
40 participants were recruited. Methods were the same as Experiment 1 with the exception of the foils: if pairs are AB and CD (with A – D being the 4 categories of sounds), the foils in the current experiment were AD and CB (cf CA and DB in Experiment 1; shown in the last column in Figure 3).

Results and Discussion

Cover task results: Participants responded correctly to the 'soft' sound with an average of 73% accuracy.

Perceptual similarity results: As in Experiments 0 and 1, an ANOVA with subject as a random factor (F_1) revealed a significant effect of category comparison (F_2 (9,351) = 46.73, $p < 0.0005$; F_1 (1,39)= 2305.6, $p < 0.0005$) with the lowest average judgments for within E1 and within E2 trials (1.45 and 1.65 respectively). Within H1 and H2 trials were on average much greater (2.30 and 2.23 respectively). Within category comparisons also yielded a significant effect of category (F_2 (3,117) = 40.44, $p < 0.0005$).

Statistical learning results: Overall, participants show ability to correctly discriminated pairs from foils (mean = 26.4, std = 4.63, $t(39) = 3.31$, $p < 0.01$), however, as in Experiment 1, performance was not uniform across Sound-Pair assignment ($F(2,39) = 5.46$, $p < 0.01$). Unlike Experiment 1, we find reliable evidence for in S-P Assign. 1 (mean = 29.54, std = 4.63, $t(12) = 4.31$, $p = 0.01$; S-P Assign.2: mean = 24.54, std = 3.23, $t(13) = 0.601$, $p > 0.5$; S-P Assign.3: mean = 25.29, std = 4.50, $t(3) = 1.07$, $p > 0.25$).



Perceptual Groupings	EE, HH	EH, EH	EH, HE	EE, HH	EH, EH	EH, HE
E1/E2/H1/H2	√	√	√	√	√	√
E1/E2/H	√	√	X	√	X	X
E/H	X	√	X	X	X	X

Figure 4: Discrimination at test for each Sound-Pair Assignment compared to predictions of whether or not we predict discrimination of pairs from foils based on both the transitional probabilities during familiarization and at test.

General Discussion

We find that adult participants can learn statistical structure using multiple exemplars from novel, complex auditory categories. Further, we demonstrate that in learning across multiple, varying exemplars, participants use the perceptual groupings available to naïve listeners rather than experimenter-defined categories.

To our knowledge, this is the first example of learning using multiple exemplars of auditory stimuli in a statistical learning paradigm. Two previous studies have used varying exemplars of visual scenes (Brady & Oliva, 2008) and human action (Loucks & Baldwin, in press). However, participants have had considerable exposure to the categories from which the variable stimuli are derived (e.g. kitchen scenes employed by Brady & Oliva, 2008), so there would be little doubt that participants would group these stimuli into the experimenter-defined categories before the experiment.

By contrast, participants in the current study have not had previous experience with the corpus of stimuli and thus have not established categorical perception of the stimuli. They do, however, have perceptual biases as determined in Experiment 0. Moreover, participants maintain their naïve perceptual groupings throughout the experiment. Both Experiments 1 and 2 demonstrate that participants rely on these groupings to learn from the transitional probabilities. Our results provide evidence that, in adults, categorical knowledge of sounds is not needed in order to learn across varying exemplars. Instead, non-categorical perceptual biases can be used to learn environmental structure (based on transitional probabilities).

To sum, we find that participants can learn based on transitional probabilities of varying exemplars for which they have no *a priori* sound categorization ability. This is a comparable task to that faced by infants in the first year of life: having to simultaneously resolve the variation in speech production and learn to segment chunks of highly coherent speech in the speech stream. We believe that the current results provide initial insight into how infants are able to learn their first words while forming their speech categories.

Acknowledgments

Thanks to Dani Rubinstein, David Kalkstein, Noa Hertzfeld, Andrew Webb, Sarah Anderson, Lori Holt, Travis Wade and Mark Seidenberg ("Statistics all the way down"), and special thanks to Michael Spivey. This work was funded by in part by R01 DC007694 and R21 DC008969.

References

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological Science, 19,* 678-685.

Graf-Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? *Psychological Science, 18,* 254 – 260.

Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6 to 12. *Journal of Phonetics, 28,* 377-396.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42.

Loucks, J. & Baldwin, D. A. (in press). When is a grasp a grasp? Characterizing some basic components of human action processing. To appear in K. Hirsh-Pasek & R. Golinkoff (Eds.), Action meets words: How children learn verbs.

Saffran, J.R., & Thiessen, E.D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), Handbook of Language Development. Cambridge: Blackwell (p. 68-86).

Wade, T. & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of Acoustical Society of America, 118,* 2618-2633.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.