Comparing the Distributional Learnability of Stops, Fricatives, Glides, and Vowels

Emily Moeng

According to the Distributional Hypothesis, infants acquire phonemes by mapping tokens into some phonetic space and noting peaks in frequencies in this space, inferring that high-frequency clusters correspond to phoneme categories (Maye et al. 2002). Although this bottom-up processing account has been influential among acquisitionists, a number of researchers have noted that natural utterances of vowels exhibit wide phonetic variability, suggesting the need for top-down sources of information, such as the infant's developing lexicon (Swingley 2009). This study seeks to quantitatively compare the distributional learnability of various types of phonemes (stops, fricatives, glides, and vowels) as spoken by a single speaker of French. An Expectation-Maximization Algorithm as implemented by the MCLUST package in R is used to estimate a purely distributional learner, and two metrics are used to evaluate the performance of this distributional learner: "Classification Accuracy" and "Psychophysical Distance Between Means." This study concludes that (1) distributional information varies widely for different phonemes, and (2) since neither metric converges in their results, more research should be done to determine which, if any, of these metrics best reflects the behavior of a human learner. Results suggest that other sources of information, such as top-down sources, must be relied upon more heavily when acquiring vowels than when acquiring stops, in order to correctly predict the order of acquisition of phonemes that infants show.

1. Introduction

Infants show language-specific discrimination of consonants around 8-12 months of age (Werker & Tees 1984, Eilers et al. 1979, Eimas et al. 1971), and language-specific discrimination of vowels even earlier, around 6-8 months of age (Kuhl et al. 1992). For example, while both English-learning and Hindi-learning infants will discriminate the retroflex /ʈa/ from the dental /t̪a/ at 6-8 months of age, around 8-12 months of age, these infants will show language-specific discriminatory abilities: the English-learning infants will no longer discriminate

---

[*] Emily Moeng, University of North Carolina at Chapel Hill, moeng@live.unc.edu.
Thank you to everyone who gave me input on this study. Special thanks to the P-side Research Group and the Acquisition Lab Group at the University of North Carolina, as well as the audiences at LSA 89, LSUGA 2, and BUCLD 40 for their helpful comments and suggestions. All errors are my own.

between these two sounds, whereas the Hindi-learning infants will continue to do so (Werker & Tees 1984). All of this occurs before infants have built a very sizable lexicon, comprehending only about 36 words by the age of 8 months, and 86 words by the age of 12 months (Caselli et al. 1995), suggesting that language-specific discrimination of contrastive phonemes cannot be dependent on infants noticing relevant minimal pairs (see also infants' failure to treat pairs of words as minimal pairs despite being given semantic evidence to the contrary; Stager & Werker 1997, Thiessen 2007).

In order to explain this early language-specific discrimination, acquisitionists have suggested a number of sources of information utilized by the language learner when forming sound categories. The most widely-known of these many sources of information is known as the Distributional Hypothesis (Maye et al. 2002, Guenther and Gjaja 1996). According to the Distributional Hypothesis, infants acquire early language-specific discrimination by mapping tokens into some *n*-dimensional phonetic space and inferring that clusters in this space correspond to sound categories, even if they do not initially end up utilizing these sound categories in lexical learning. Other sources of information include the role of semantics (Hayes-Harb 2007) and the role of overall lexical shape, devoid of any semantic information (Feldman et al. 2013).

The goal of this study is to quantitatively compare the distributional learnability of stops (/k g/), fricatives (/s ʃ/), glides (/j ɥ w/), and vowels (/i y u/) within a single speaker of French, in order to determine if distributional learning alone is sufficient to correctly predict the order of acquisition of phonemes that human learners show.

2. Background

2.1. The Distributional Hypothesis

According to the Distributional Hypothesis, a language learner maps tokens into some phonetic space, and uses their relative frequencies to infer the number of sound categories (s)he is being exposed to. Language learners exposed to a bimodal distribution of tokens along some phonetic dimension(s) will infer that there are two sound categories, whereas language learners exposed to a monomodal distribution will infer that there is only one. Therefore, in a purely distributional-based learning account, the language learner's job is to 1) extract sounds from a stream of speech, 2) map out each extracted sound into some phonetically-based *n*-dimensional space, 3) keep track of past sounds (s)he has mapped within this space, and 4) infer that each region of high frequency tokens represents a phoneme in the ambient language. Artificial learning tasks show that language learners are capable of these computations. Maye et al. (2002) find that infants exposed to a simplified artificial language with a bimodal distribution of tokens ranging between a voiceless unaspirated stop [t] and a pre-voiced stop [d] inferred that there were two categories, whereas a group exposed to a language with a monomodal distribution inferred that there was only one category. This finding has been replicated for the stimuli [g] – [k], [a] – [ɑ], [i] – [ɪ], and [ç] – [ʂ] (Hayes-Harb 2007, Gulian et al. 2007, Cristià et al. 2011).

2.2. Distributional Learning is only one of many sources of information for language learners

While these artificial learning experiments show that language learners are able to make the statistical calculations required for the Distributional Hypothesis, a number of researchers have recently pointed out that the distributions in natural language exhibit too much phonetic overlap for distribution to be the *only* source of information that language learners rely upon. Hayes-Harb (2007) shows that language learners utilize semantic information when determining boundaries between phonemes (e.g. participants who are shown that [ga] means 'pot', but [ka] means 'mouse' are more likely to treat [g] and [k] as different phonemes than participants who are shown [ga] and [ka] both mean 'pot'). Feldman et al. (2013) shows that overall lexical shape, devoid of any semantic information, also affects a language learner's acquisition of phoneme categories (e.g. participants who hear [a] and [ɔ] in similar lexical contexts [lit**a**] and [lit**ɔ**] are more likely to treat [a] and [ɔ] as single phonemes than participants who hear them in different lexical contexts [lit**a**] and [gut**ɔ**]). Each of these is just one of many sources of information which a language learner makes use of when carrying out the difficult task of categorizing incoming information, and each may be weighted differently in learners' minds. For example, Hayes-Harb (2007) compares the effect of semantic information with the effect of distributional information on the formation of phonemes, finding semantic information to have a stronger effect than distributional information, and Moeng (2015) compares the effect of lexical shape to distributional information, although results were inconclusive.

This paper will reference both "sources of information" and "(acoustic) cues." These concepts are similar in that they both acknowledge that learners often take many factors into account when creating and utilizing categories, but each applies to different steps of the categorization process. "Source of information" will refer to the factors taken into account by the learner when (s)he is in the process of creating and acquiring phoneme categories. One of these sources of information is Distributional Learning. Other sources are the earlier-referenced semantic information and lexical shape information. "(Acoustic) cue" will refer to some acoustic aspect of a sound which is taken into account when language users are deciding which already-existing category or phoneme a sound belongs to, such as Voice Onset Time, or formant values.

2.3. Are different phonemes acquired differently?

Because language learners utilize multiple sources of information when acquiring sound categories, there is the chance that different types of phonemes (e.g. stops, fricatives, glides, and vowels) are acquired with a heavier or lighter reliance upon each of these sources of information. For example, if distribution is weighted (i.e. relied upon) *x* amount by an infant acquiring phonemes, and other sources of information (semantics, lexical shape, etc.) are weighted *y* amount by the infant, are these weights kept constant for all types of phonemes? Is there reason to believe that language learners do not use the same blanket approach in acquiring all types of phonemes?

This question is worth pursuing, as some proposals predict that language learners process different types of phonemes in different ways. For example, proponents of what is known as the CV Hypothesis argue that there is a categorical distinction between the processing of consonants and vowels (Nespor et al. 2003, Bonatti et al. 2005, Toro et al. 2008, Havy & Nazzi 2009), rather than two ends of a sonority spectrum or a convenient name for a bundle of features (Caramazza et al. 2000). Ramus et al. (1999) and Mehler et al. (1996) suggest that infants treat vowels as speech but consonants as “unanalyzed noise.”

Another reason to compare the distributional strength of different types of phonemes is that arguments which have been made against the Distributional Hypothesis point to the great amount of category overlap that vowels exhibit, arguing that this phonetic overlap is so high that, without other sources of information, a language learner could not possibly pick out categories based on frequency distributions alone. For example, Figure 1 shows 11 English monophthongs. According to a distributional account, 11 clusters should be visible, and yet this does not seem to be the case due to the wide phonetic overlap between each vowel.

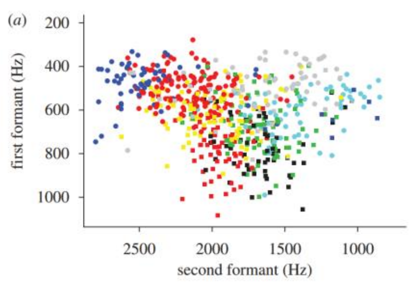

Figure 1. The problem of overlapping categories for a purely distributional account. Figure from Swingley (2009) maps out 11 English monophthongs. According to a distributional account, 11 clusters should be visible.

However, this argument against the distributional account seems to have only been made for vowels. A study by Lisker and Abramson (1964) suggests that stops differing in VOT do not pose the same problem as vowels do to distributional learning, as stop VOTs do not seem to overlap to the extent that vowel formants do. As seen in Figure 2, English speakers are exposed to a distribution with two prominent peaks and one smaller peak. We could imagine either a model in which language learners only notice the more prominent peaks in frequency (that is, peaks in frequency that fall above some threshold), or a model in which language learners notice all local maxima in frequency (3 in this case), and then, through some second step, collapse phoneme categories which are in complementary distribution into a single category. By comparison, Dutch speakers (Figure 2, right), who have a single phoneme associated with velar stops, are exposed to only a single peak in frequency.

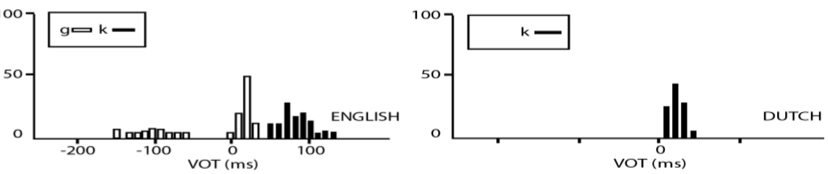

Figure 2. In a figure plotting VOT for velar oral stops, we see two or three clear peaks form in the English-speaking data (left), and only one peak form in the Dutch data. Figures adapted from Lisker and Abramson (1964).

This is precisely the type of distribution a purely distributional hypothesis would expect: a language with two phonemes exhibits two (or three) frequency peaks, while a language with a single phoneme exhibits only a single peak in frequency.

While a visual comparison of Figures 1 and 2 suggests that stops would be better learned by a distributional learner than vowels, this study seeks to make a quantitative comparison of stops and vowels, along with fricatives and glides, by evaluating the results of a Gaussian fitting model. In addition, the measurements made to create Figures 1 and 2 were taken from different studies and different elicitation methods. This study will compare various types of phonemes, all of which are produced by a single speaker in a single language in the same discourse style, in order to most accurately compare the distributions of these phonemes.

3. Research Questions

The goal of this paper is to answer the question: *How do the distributions of stops, fricatives, glides, and vowels compare?*

To answer this question, natural utterances of stops /k g/, fricatives /s ʃ/, glides /j ɥ w/, and vowels /i y u/ were measured in French child-directed speech. Each is a phoneme in French, so a distributional learner would expect each to exhibit a clear peak in frequency.

4. Methods

Because this study focuses on the type of input the infant receives, measurements are made on child-directed speech. Phonetic cues were measured from the speech of one male in his early 30’s speaking European French from the Yamaguchi corpus in CHILDES (Yamaguchi 2007, MacWhinney 2000). The Yamaguchi corpus contains naturalistic dialogue of a mother and father interacting with their infant, recorded at roughly one-month intervals across the span of 3 years. Files were chosen from approximately halfway through this 3-year span in hopes that the speakers were by then comfortable with being recorded and were acting (and speaking) naturally.

1-2 salient phonetic cues were measured for each phoneme type: voice onset time (VOT) for stops, center of gravity (CoG) for fricatives, and the second and third formants (F2 and F3) for vowels and glides.

When measuring voice onset time, the onset of voicing was determined to be the point at which the waveform began to exhibit clear periodic behavior, and the stop release was marked at the point where intensity increased. Stops in codas were not included in measurements.

Regarding vowel and glide measurements, vowel formants were measured at the midpoint of the vowel. Glides were marked where there were clear visible formants in the spectrogram, and a low-intensity, periodic waveform. In order to avoid as much coarticulation from neighboring vowels as possible, glide formant measurements were made at the farthest point from the glide's neighboring vowel(s), which differed based on the glide's phonological environment: formants for glides were measured 10% of the way into the phoneme if the glide followed a consonant and preceded a vowel, 90% of the way in if it followed a vowel and preceded a consonant, and 50% of the way in if it was intervocalic.

All measurements were made using the speech analysis software Praat (Boersma & Weenink 2001). Identities of underlying phonemes were based on phonemic transcriptions (marked as "%mod" within CHILDES) as marked by the transcriber of the Yamaguchi corpus. A total of 2007 phonemes were measured.

It should be noted that the choice to measure only one or two salient cues does not mean that it is believed that any of these cues alone will be sufficient for a learner. As noted earlier, there are likely many sources of information that the language learner must draw upon when determining phonemes belonging to his or her language, only one of which is distributional in nature. The purpose of this study is not to claim that distribution alone is sufficient, but rather to determine the relative strengths of distributional information across different types of phonemes within a single language.

5. Analysis

Section 5.1 presents the categories learned by the machine learner next to the actual phoneme categories. Section 5.2 evaluates how well the leaner learned each of these categories using two metrics introduced here: Classification Accuracy and Psychophysical Distance Between Means.

5.1. Learned Gaussian distributions of stops, fricatives, glides, and vowels

Gaussian distributions for each phoneme were estimated with the Expectation-Maximization (EM) algorithm as utilized by the MCLUST package in R (R Core Team 2015, Fraley et al. 2012, Fraley & Raftery 2002). This unsupervised machine learner was given no information as to the number of categories to be learned, and discovered parameters such as orientation, variance, and mean without any outside input.

For the stops, the machine learner was given only one dimension: VOT. Actual and learned stop categories are shown in Figure 3. The machine learner discovered three categories when there were only two phonemes. Because each of the three learned categories contained quite a few tokens, this study looked more closely at the members within each category and their phonological environment. It was found that many of the phonemes in the third category were /k/ followed by a high vowel or glide. As /k/ has been shown to have greater VOT following high tense vowels (Klatt 1975, Port & Rotunno 1979), this study looked at /k/ following high vowels/glides ("k_H," dark green) separately from the elsewhere-/k/ (/k/ everywhere other than before a high vowel/glide, light green). Note that the discovery of too many categories is not considered problematic for a distributional account, as the discovery of complementary distribution can later collapse these into a single phoneme category (see the discussion concerning Figure 2).

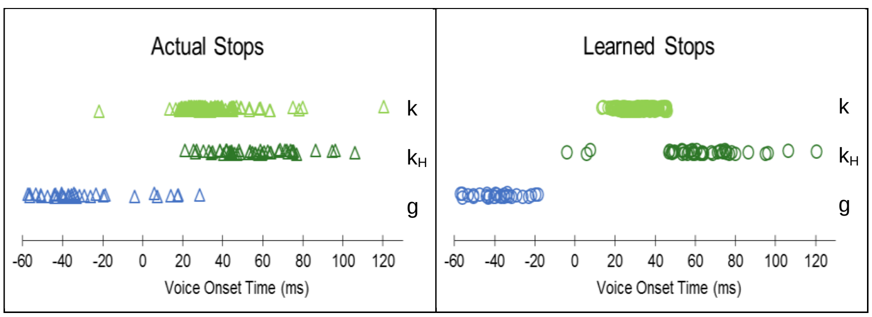

Figure 3. Actual phonemes (left) and learned categories (right) for the stops /k/, /k/ preceding high vowels and glides ('k_H'), and /g/.

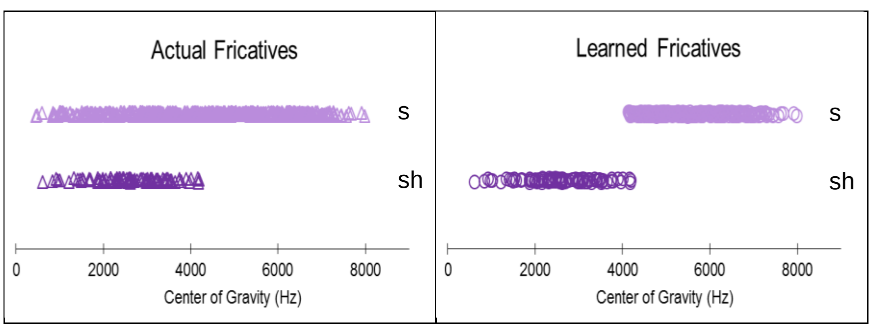

Figure 4. Actual phonemes (left) and learned categories (right) for the fricatives /s/ and /ʃ/.

For the fricatives, the machine learner was only given center of gravity measurements. Actual and learned fricative categories are given in Figure 4. The machine learner correctly discovered two categories: one corresponding to /s/ (light purple) and one corresponding to /ʃ/ (dark purple).

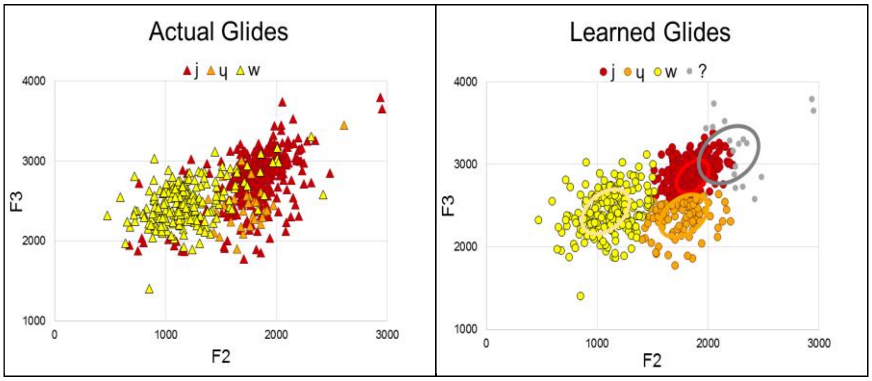

Figure 5. Actual phonemes (left) and learned categories (right) for glides.

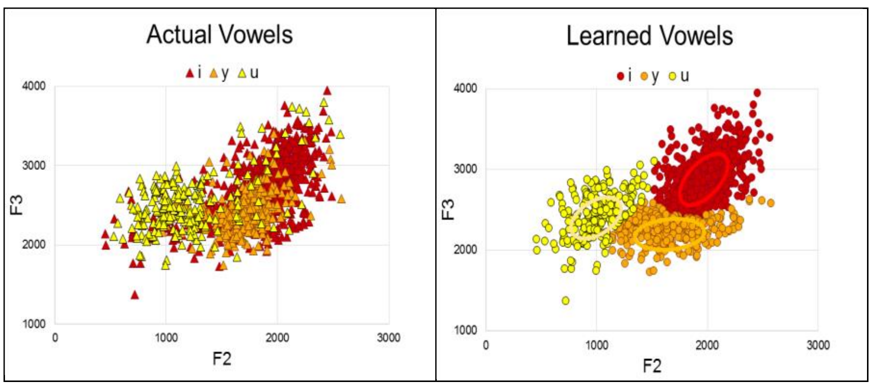

Figure 6. Actual phonemes (left) and learned categories (right) for vowels.

For vowels and glides, the distributional machine learner was given two dimensions: F2 and F3. Actual and learned categories of vowels and glides are shown in Figures 5 and 6. The distributional learner correctly discovered three vowel categories, as indicated in the bottom right panel of Figure 6. However, the learner discovered an extra (sparsely-populated) fourth category when given the glide distributions, as shown in Figure 5.

5.2. Evaluating the distributional learner's performance

This section evaluates the "learnability" of each of these phonemes by assessing how accurate the previous section's learned Gaussian categories are. The learnability of each phoneme was judged along two metrics: (1) by calculating what will be referred to here as Classification Accuracy, the percent of phonemes that were correctly classified by the machine learner (roughly following the methodology of Adriaans & Swingley (2012)), and (2) by calculating what will be referred to here as Psychophysical Distance Between Means, the perceptual distance between the actual mean of each phoneme category and the learned Gaussian category mean. The justification behind Classification Accuracy as a metric of learnability is that, the better a category has been learned, the more accurately it will correctly classify actual tokens into the learned category. The justification behind Psychophysical Distance Between Means as a metric of learnability is that, the better a category has been learned, the closer the learned "prototype" or category mean will be to the actual category mean.

Table 1. Classification Accuracy for stops, fricatives, glides, and vowels (percent of phonemes correctly classified by the machine learner).

Type	Phoneme	Classification Accuracy (% Correctly Classified)
Stop	/k/ elsewhere	91% (138/152)
	/k/ before [high]	59% (30/51)
	/k/ overall	83% ((138+30)/(152+51))
	/g/	80% (28/35)
Fricative	/s/	58% (355/615)
	/ʃ/	97% (95/98)
Glide	/j/	71% (257/361)
	/ɥ/	75% (24/32)
	/w/	85% (167/196)
Vowel	/i/	73% (477/651)
	/y/	58% (176/304)
	/u/	68% (154/225)

Classification Accuracy, the percent of phonemes correctly classified, is indicated in Table 1. For example, 80% of all /g/s measured were correctly placed into the learned "g" category by the machine learner. According to this metric, the /y/, /u/, /ʃ/, and /k/ before high vowels performed poorly, being placed into the correct category only 58%, 68%, 58%, and 59% of the time, respectively (see Adriaans and Swingley (2012) for similar numbers regarding a metric similar to Classification Accuracy for English vowels). All other phonemes (/k g s j ɥ w i/) were placed into the correct learned category 71% - 91% of the time, with the "elsewhere-/k/" and /g/ performing the best, being correctly classified 91% and 80% of the time, respectively.

The second metric for learnability, Psychophysical Distance Between Means, is designed to measure how well the learner learned the acoustic location of the prototypical phoneme category. This involved measuring the distance between the actual mean of the phonetic variable given to the machine learner (VOT for stops, CoG for fricatives, F2 and F3 for glides and vowels), and the learned mean of that same phonetic variable according to the Gaussian distribution discovered by the machine learner. In order to compare acoustic distance between an actual and learned VOT to an actual and learned formant value, acoustic distances were converted into psychophysical, or perceptual, distances. To do this, acoustic distances between the actual and learned averages were divided by the length of 1 Just-Noticeable Difference (JND).

Table 2. Psychophysical Distance Between Means between the actual and learned category center for stops, fricatives, glides, and vowels.

Type	Phoneme	Actual Category Mean (ms, Hz)	Learned Category Mean (ms, Hz)	|Actual-Learned| (ms, Hz)	Psychophysical Dist. Btwn. Means |Actual-Learned| (JNDs)
Stop	/k/	33 ms (VOT)	31 ms (VOT)	2 ms	0.09 – 0.13 JND (2/23), (2/16)
	/k/-high	55 ms (VOT)	62 ms (VOT)	7 ms	0.30 – 0.44 JND (7/23), (7/16)
	/g/	-29 ms (VOT)	-39 ms (VOT)	10 ms	0.43 – 0.63 JND (10/23), (10/16)
Glide	/j/	1817 Hz (F2)	1847 Hz (F2)	30 Hz	1.1 JND (30/28)
		2785 Hz (F3)	2889 Hz (F3)	104 Hz	---
	/ɥ/	1713 Hz (F2)	1752 (F2)	39 Hz	1.5 JND (39/26)
		2429 Hz (F3)	2437 Hz (F3)	8 Hz	---
	/w/	1191 Hz (F2)	1114 (F2)	77 Hz	4.5 JND (77/17)
		2479 Hz (F3)	2436 Hz (F3)	43 Hz	---
Fricative	/s/	4347 Hz (CoG)	5583 Hz (CoG)	1236 Hz	---
	/ʃ/	2604 Hz (CoG)	2619 Hz (CoG)	15 Hz	---
Vowel	/i/	1883 Hz (F2)	1956 Hz (F2)	73 Hz	2.5 JND (73/29)
		2771 Hz (F3)	2858 Hz (F3)	87 Hz	---
	/y/	1704 Hz (F2)	1654 Hz (F2)	50 Hz	2.0 JND (50/25)
		2423 Hz (F3)	2296 Hz (F3)	127 Hz	---
	/u/	1205 Hz (F2)	1032 Hz (F2)	173 Hz	12 JND (173/15)
		2529 Hz (F3)	2456 Hz (F30	73 Hz	---

JNDs were taken to be 16-23 ms for VOTs[†] (16 ms for across-category discrimination, 23 ms for within-category discrimination; values from the *bilabial* stops [p]-[b] from Hazan et al. (2009: Table 1)), and 1.5% of the learned formant average for F2 (Kewley-Port & Watson 1994). Psychophysical Distances Between Means were not obtained for fricatives or for F3 values for vowels and glides, as JND values were not found by the author.

According to this metric, learners perform more poorly the greater the Psychophysical Distance Between Means. Learners can be considered to perform well if Psychophysical Distance Between Means falls below the level of discrimination (i.e. under 1 JND). As can be seen in Table 2, stops performed well according to this metric[‡], while glides and vowels did not.

Table 3. Summary of performance of the machine learner. The top of the table indicates phonemes which were learned well by the learner, according to each metric.

Classification Accuracy (↓)	Type	Phoneme
97%	Fricative	/ʃ/
91%	Stop	/k/
85%	Glide	/w/
80%	Stop	/g/
75%	Glide	/ɥ/
73%	Vowel	/i/
71%	Glide	/j/
68%	Vowel	/u/
59%	Stop	/k/-high
58%	Fricative	/s/
58%	Vowel	/y/

Psychophysical Dist. Btwn. Means (↑)	Type	Phoneme
0.06	Stop	/k/
0.2	Stop	/k/-high
0.29	Stop	/g/
1.1	Glide	/j/
1.5	Glide	/ɥ/
2	Vowel	/y/
2.5	Vowel	/i/
4.5	Glide	/w/
12	Vowel	/u/

Table 3 summarizes the performance of the distributional learner as measured by each of these two metrics, with phonemes which were learned well at the top of the table (as indicated with higher numbers for Classification Accuracy and lower numbers for Psychophysical Distance Between Means) and phonemes which were learned poorly at the bottom of the table. It should be noted that, although these two metrics were designed to measure the same thing (that is, learner performance) these two metrics did not always converge. For example,

[†] Because stops exhibit categorical perception along the VOT dimension, with within-category members being perceived as being more similar to one another than across-category members (Eimas et al. 1971), both within-category and across-category JND values were used, resulting in a range of possible Psychophysical Distances Between Means.

[‡] For both JND values derived from within-category members and JND values derived from across-category members, Psychophysical Distance Between Means was below 1 JND, and therefore should not be a perceivable difference of category means.

while /k/ before high vowels performed well according to the metric of Psychophysical Distance Between Means, it performed poorly according to Classification Accuracy.

6. Conclusion

By looking at different types of phonemes drawn from the same set of utterances, this study attempted to compare distributional qualities of stops, fricatives, glides, and vowels taken from natural child-directed speech. However, it was found that the two metrics used to measure the performance of the distributional learner did not converge on a single ranking, with Psychophysical Distance Between Means conforming more to the intuition that stops exhibit distributions that are highly conducive for a distributional learner, as we might imagine given Figure 2. Because the Distributional Hypothesis is such an influential proposal, it is believed the field will benefit from further work to determine how to best measure distributional learner performance.

With that caveat in mind, it generally seemed to be the case that stops performed more poorly than vowels according to both metrics. Before detailing the implications of this finding, it should be noted that this study gave vowels a distributional advantage over stops, by giving the machine learner two acoustic dimensions to work in for vowels (F2 and F3), and yet only a single dimension for the stops (VOT). Because there is the possibility for greater and more accurate category distinction when given more acoustic dimensions on which to differentiate two given tokens, feeding multiple cues to the machine learner should have strengthened the distributional learnability of vowels. And yet, even when given this multiple-cue advantage over the stops, vowels still performed worse than the stops.

The above observation that vowels exhibit weak distributional information (at least, with F2 and F3 as its dimensions) is significant because, unless the wrong phonetic cues were chosen for this study, if we again assume that distribution is weighted *x* amount and other sources of information (semantics, lexical shape, etc.) are weighted *y* amount, for all phoneme categories, we would expect vowels, with their weaker distributions to be acquired *later* than other phoneme types. And yet when we observe actual infants, vowels exhibit language-specific discrimination earlier (around 6-8 months) than consonants. To explain both the earlier language-specific discrimination of vowels as well as the poorer performance of vowels in terms of distribution, other sources of information must more than make up for the lack of distributional strength in vowels. In other words, it must be the case that other sources of information are more heavily weighted (i.e. relied upon more) for stops than for vowels. This paper motivates further investigation of the relative roles of different sources of information across more phoneme types, since this may have implications regarding the methods learners utilize when acquiring different phoneme types.

References

Adriaans, Frans & Daniel Swingley. 2012. Distributional learning of vowel categories is supported by prosody in infant-directed speech. In Miyake, Peebles, & Cooper (eds.), Annual Conference of the Cognitive Science Society 34. Austin, TX: Cognitive Science Society.

Boersma, Paul & David Weenink. 2001. Praat, a system for doing phonetics by computer. 341-345.

Bonatti, Luca, Marcela Peña, Marina Nespor & Jacques Mehler. 2005. Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. Psychological Science 16(6). 451-459.

Caramazza, Alfonso, Doriana Chialant, Rita Capasso & Gabriele Miceli. 2000. Separable processing of consonants and vowels. Nature 403(6768). 428-430.

Caselli, Maria Cristina, Elizabeth Bates, Paola Casadio, Judi Fenson, Larry Fenson, Lisa Sanderl & Judy Weir. 1995. A cross-linguistic study of early lexical development. Cognitive Development 10(2). 159-199.

Cristià, Alejandrina, Grant L. McGuire, Amanda Seidl & Alexander L. Francis. 2011. Effects of the distribution of acoustic cues on infants' perception of sibilants. Journal of Phonetics 39(3). 388-402.

Eilers, Rebecca E., William Gavin & Wesley R. Wilson. 1979. Linguistic experience and phonemic perception in infancy: A crosslinguistic study. Child Development 50(1). 14-18.

Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk & James Vigorito. 1971. Speech perception in infants. Science 171(3968). 303-306.

Feldman, Naomi, Emily Myers, Katherine White, Thomas L. Griffiths & James L. Morgan. 2013. Word-level information influences phonetic learning in adults and infants. Cognition 127(3). 427-438.

Fox, John & Sanford Weisberg. 2011. An {R} Companion to applied regression, 2nd edn. Thousand Oaks, CA: Sage.

Fraley, Chris & Adrian E. Raftery. 2002. Model-based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association 97. 611-631.

Guenther, Frank H. & Marin N. Gjaja. 1996. The perceptual magnet effect as an emergent property of neural map formation. The Journal of the Acoustical Society of America 100(2). 1111-1121.

Gulian, Margarita, Paola Escudero & Paul Boersma. 2007. Supervision hampers distributional learning of vowel contrasts. International Congress of Phonetic Sciences 16. 1893-1896.

Havy, Mélanie & Thierry Nazzi. 2009. Better processing of consonantal over vocalic information in word learning at 16 months of age. Infancy 14(4). 439-456.

Hayes-Harb, Rachel. 2007. Lexical and statistical evidence in the acquisition of second language phonemes. Second Language Research 23(1). 65-94.

Hazan, Valerie, Souhila Messaoud-Galusi, Stuart Rosen, Suzan Nouwens & Bethanie Shakespeare. 2009. Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit? Journal of Speech, Language, and Hearing Research 52(6).

Kewley-Port, Diane & Charles S. Watson. 1994. Formant-frequency discrimination for isolated English vowels. The Journal of the Acoustical Society of America 95(1).

Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech, Language, and Hearing Research 18(4). 686-706.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255(5044). 606-608.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20(3). 384-422.

MacWhinney, Brian. 2000. The CHILDES Project: Tools for analyzing talk, 3rd edn., vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.

Maye, Jessica, Janet F. Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82(3). B101-B111.

Mehler, Jacques, Emmanuel Dupoux, Thierry Nazzi & Ghislaine Dehaene-Lambertz. 1996. Coping with linguistic diversity: The infant's viewpoint. In Morgan, James L. & Katherine Demuth (eds.), Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. 101-116.

Moeng, Emily. 2015. Acquiring phonemes: Is frequency or the lexicon the primary cue? Penn Linguistics Colloquium (PLC) 38.

Nespor, Marina, Marcela Peña & Jacques Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. Lingue e Linguaggio 2(2). 203-229.

Peperkamp, Sharon, Michéle Pettinato & Emmanuel Dupoux. 2003. Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (eds.), Boston University Conference on Language Development (BUCLD) 27. 650-661. Sommerville, MA: Cascadilla Press.

Port, Robert F. & Rosemarie Rotunno. 1979. Relation between voice-onset time and vowel duration. The Journal of the Acoustical Society of America 66(3). 654-662.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org.

Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73(3). 265-292.

Stager, Christine L. & Janet F. Werker. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640). 381-382.

Swingley, Daniel. 2009. Contributions of infant word learning to language development. Philosophical Transactions of the Royal Society B: Biological Sciences 364(1536). 3617-3632.

Thiessen, Erik D. 2007. The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language 56(1). 16-34.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7(1). 49-63.

Yamaguchi, Naomi. 2007. Markedness, Frequency: Can We Predict the Order of Acquisition of Consonants? Second Oxford Postgraduate Conference in Linguistics. 236-243.