Effects of the distribution of acoustic cues on infants' perception of sibilants

Alejandrina Cristià[*,a], Grant L. McGuire[c], Amanda Seidl[b], and Alexander L. Francis[b]

[*] Corresponding author: alecristia@gmail.com, Tel: 33-662-96-0572

[a] Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS-ENS-DEC-CNRS, Paris, 75005, France

[b] Purdue University, West Lafayette, 47901 Indiana, USA

[c] University of California at Santa Cruz, Santa Cruz, 95064 California, USA

Abstract

Infants may converge on their native speech categories by attending to frequency distributions that occur in the acoustic input. To date, the only empirical support for this statistical learning hypothesis comes from studies where a *single salient* dimension was manipulated. In this paper, the statistical hypothesis is pushed one step further, by introducing multidimensionality, and using a less salient pair of sounds. English-learning infants were exposed to a bidimensional continuum between retroflex and alveopalatal sibilants. In the first experiment, infants heard one of two distributions (no peaks, or two peaks), and were tested with sounds varying along only one dimension. Infants' responses differed depending on the familiarization distribution, and their performance was equally good for the vocalic and the frication dimension, lending some support to the statistical learning hypothesis. However, this learning was restricted to the retroflex category. In a second experiment, infants heard a highly kurtotic unimodal distribution centered in the previously unlearned alveopalatal category. Again, infants failed to make any distinction within this region. In summary, these results contribute limited support for the statistical hypothesis by extending findings to a multidimensional space.

Keywords

Statistical learning; place of articulation; fricatives

Effects of the distribution of acoustic cues on infants' perception of sibilants

1.0 Introduction

Research on the development of speech development suggests that, in early infancy, humans show speech discrimination abilities that do not appear to depend on their language experience, while by the end of their first year, their perception is more tuned to the sounds present in the ambient language (Jusczyk, 1997). In this paper, we present evidence providing moderate support to the hypothesis that, if this tuning is due to category learning, then it may be explained as a result of attention to statistical distributions of acoustic cues in the speech infants hear.

More specifically, we first review, in §1.1, previous results on infants' perceptual acquisition that suggest that this process involves the formation of categories on the basis of pre-existing auditory-perceptual abilities (as proposed by e.g., Aslin & Pisoni, 1980, and Kuhl, 2009), and not simply the selection of a subset of categories among an innately given set (as suggested by e.g., Eimas, 1975, Gervain & Werker, 2008, among others). If perceptual acquisition involves learning, it is plausible that this process is aided by infants' attention to statistical distributions of acoustic cues. This *statistical learning* hypothesis and extant empirical evidence supporting it is summarized in §1.2. Nonetheless, the stimuli used in previous research relied on a single, psychoacoustically salient dimension. Therefore, there is still little evidence that the statistical learning hypothesis can scale up to the challenges infants face in the task of natural language acquisition, as argued in §1.3. This evidence was sought in two experiments, whose motivation and design is introduced in §1.4, and reported on §2 and 3. These studies show that infants' perception is affected by acoustic cue distributions, but possibly only in regions in acoustic space to which infants are already sensitive. The implications of these and other findings are discussed in §4.

1.1 Infants' perception: Learning or selection?

It is commonly reported that infants are able to discriminate contrasts that do not exist in their ambient language. For example, Japanese 6-month-old infants can discriminate the non-native contrast /r-l/ (e.g., Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006), a contrast that is remarkably difficult for their elders to hear (e.g., Iverson, Kuhl, Akahane-Yamada, Diesch, Tokhura, Kettermann, & Siebert, 2003). However, it is equally important to remember that early discrimination abilities are not unbounded. Since we return to the question of initial sensitivies below, we give just two examples here: both English- and French-learning 6- to 8-month-olds perform poorly with /d-ð/ (Polka, Colantonio, & Sundara, 2001), and neither Filipino- nor English-hearing 6- to 8-month-olds can discriminate /n-ŋ/ (Narayan, Werker, & Speeter Beddor, 2009).

A second fact of infant perception is that, by about 12 months of age, monolingual infants' sensitivity tends to improve (or remain the same) for contrasts *present* in their ambient language, and to decline for others that are *not functional* in that language (e.g., Cheour, Alho, Ceponiene, Reinikainen, Sinio, Pohjavouri, Aaltonen, & Naatanen, 1998; Kuhl et al., 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Mattock & Burnham, 2006; Mattock et al., 2008; Polka & Werker, 1994; Seidl, Cristià, Bernard, & Onishi, 2009; Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk, & Best, 1993; Werker et al., 1981; Werker & Tees, 1984). For example, at 12 months, Filipino learners succeed with /n-ŋ/, while English learners fail (Narayan et al., 2009), but English-hearing infants are much better at discriminating /r-l/ than Japanese-hearing children (Kuhl et al., 2006). However, two caveats should be made. First, note that patterns of decline appear to be modulated by the frequency of the sounds (Morgan, Anderson, & White, 2004); and by their articulatory/perceptual characteristics (Best & McRoberts, 2003; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Best, McRoberts, & Sithole, 1988). Second, in some cases, infants' sensitivity remains *equally bad* by the end of the first year. This is the case for /d-ð/ (Polka et al., 2001), and /s-S/ (Nittrouer, 2001). We return to this question below.

Since infants are unable to discriminate some contrasts in the absence of experience (e.g., /n-ŋ/), and given that performance later in development is also modulated by experience (e.g., frequency, acoustic salience), these findings provide little support for theories that reduce phonetic acquisition to the selection of features present in one's ambient language among an innately given feature set (Eimas, 1975; Liberman & Mattingly, 1985). Instead, Aslin & Pisoni (1980) propose that the developmental change in infant perception could involve four category learning processes, as follows. When discrimination abilities are robust in the absence of experience, subsequent exposure to a language having that contrast will result in (1) maintenance, whereas exposure to a language lacking that contrast could result in (2) attenuation. In contrast, when discrimination abilities are initially weak, exposure to a language that recruits that contrast should result in (3) enhancement, or possibly (4) induction, in order to achieve native perception. Given that much of this learning appears to take place before infants acquire a sizable lexicon (Caselli, Bates, Casadio, Fenson, Fenson, Sanderl et al., 1995), it cannot be lexical knowledge that drives perceptual acquisition. On the contrary, the mechanism involved in perceptual acquisition must not only be able to account for the abovementioned 4 processes, but also do it on the basis of exposure to speech, without resource to lexical information.

1.2 Statistical learning

Recent research in other domains of language acquisition underlines infants' astounding abilities to learn from statistical patterns (see, for instance, Aslin & Newport, 2009; Saffran, 2009, for recent reviews). A version of this *statistical learning* hypothesis is developed for phonetic acquisition by Pierrehumbert (2003). There, it is proposed that infants may initiate (rudimentary) categories by keeping track of differential frequency distributions represented in their spoken input. Overlaying this description onto the 4 processes detailed above, we can propose that cases of maintenance, enhancement and decline are explained as follows: infants cease to attend to initially discriminable contrasts that do not coincide with modes in frequency, and preserve their sensitivity to those that are represented by modes in frequency in perceptual space. Induction, in contrast, is more difficult to explain within this description. That is, if there is absolutely no sensitivity for a certain acoustic distinction, exposure to statistical distributions that span this region in acoustic space are not informative, as they fall on the same region of perceptual space. We defer a fuller discussion of induction to SS4.3, and conclude for the time being that statistical learning could easily capture at least 3 processes involved in phonetic acquisition: maintenance and decline of initially robust sensitivities, and enhancement of initially weak ones.

Recently, 2 of those 3 processes have been replicated in the laboratory (Maye, Weiss, & Aslin, 2008; Maye, Werker, & Gerken, 2002). In those studies, Maye and colleagues investigated the effect of short exposures to different distributions of voice onset time (VOT) on infants' perception of voiceless and aspirated stops (Maye, Werker, & Gerken, 2002) and prevoiced and voiceless stops (Maye, et al., 2008). Infants heard a small number of tokens varying in VOT for a few minutes. For infants hearing a bimodal distribution, tokens with VOT values close to the ends of the continuum were especially frequent. Contrastingly, for infants in the unimodal condition, tokens with VOT values in the center of the continuum occurred most frequently. Results suggested that having heard a bimodal distribution allowed infants to discriminate between tokens near the ends of the distribution, while a unimodal distribution was as inefficient in affecting infants' perception as an initial exposure to irrelevant tones (Maye, et al., 2008, unimodal versus control).

1.3 Limitations of extant research

While these results are encouraging, the learning situations infants faced in those studies do not exhaust the challenges encountered in natural language acquisition, which include reduced cue salience and multidimensionality. These are detailed in the following subsections.

1.3.1 Salience

As mentioned above, both previous studies have manipulated VOT. Much research shows that VOT has rather unique psychoacoustic properties, which impact performance with contrasts involving this dimension. With respect to the psychoacoustic properties of VOT, infants with little experience exhibit categorical perception of the voiceless-aspirated contrast, dishabituating to an acoustic change that spanned the voiceless-aspirated boundary, but not when the same acoustic distance was spanned within one of those categories (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976). Some evidence also suggests that infants can discriminate prevoiced from voiceless tokens in the absence of significant experience (Aslin, Pisoni, Hennessy, & Perey, 1981; Eimas, 1975; but see Lasky et al., 1975, for arguments that this contrast may be harder for infants than the voiceless-aspirated one). Furthermore, non-human animals also display categorical perception of the voiceless-aspirated contrast (e.g., Kuhl & Miller, 1975, 1978; Kuhl & Padden, 1982). Finally, categorical perception phenomena of both VOT contrasts (prevoiced-voiceless and voiceless-aspirated) has been replicated with non-speech continua in both adults (e.g., Elangovan & Stuart, 2008; Pisoni, 1977) and infants (Jusczyk, Pisoni, Walley, & Murray, 1980; Jusczyk Rosner, Reed, & Kennedy, 1989). These 3 facts strongly suggest that, for some regions of acoustic space, there are discontinuities along the VOT dimension that are psychoacoustic in nature.

As for learnability, at least two lines of research suggest that these psychoacoustic characteristics may boost the learnability of voicing categories. First, improved performance has been documented for contrasts involving this dimension (e.g., within-category discrimination training, Pisoni & Lazarus, 1974; discrimination under cognitive load, Gordon, Eberhardt, & Rueckl, 1993). Secondly, both speech and non-speech training studies with adults support the hypothesis that VOT discontinuities facilitate category learning specifically. For example, a few minutes of training suffice to re-train American listeners on the prevoiced-voiceless contrast, which maps onto a single phonemic category in their ambient language (Pisoni, Aslin, Perey, & Hennessy, 1982; see also Tees & Werker, 1984, who document that voicing contrasts are easier to re-learn than place contrasts). In addition, laboratory learning of a category centered in the VOT auditory discontinuities is remarkably difficult, and much more so that learning of a category that does not span this region (Holt, Lotto, & Diehl, 2004). Thus, these findings are compatible with the idea that the psychoacoustic properties of VOT draw attention and constrain learning, facilitating the acquisition of categories whose boundaries coincide with the documented auditory discontinuities along that dimension.

Unfortunately for learners, few phonetic contrasts depend on this salient VOT dimension, and research should be carried out to document perceptual changes along other dimensions. Furthermore, one may argue that previous studies have only explored cases of maintenance and decline, and they provide little evidence for the cases involving initially weak contrasts. To provide such evidence, it is necessary to select a set of sounds differing along a dimension on which such auditory-based discontinuities have not been documented. In this quest, the place of articulation contrasts /b-d/ and /b-g/ in stops and /w-j/ in glides, the manner contrasts /b-w/, /b-m/, and /r-l/ are possibly not good candidates, as all of these are easily discriminable by very young infants (Bertoncini et al., 1987; Eimas, 1975a; Eimas & Miller, 1980a, 1980b; Miller & Eimas, 1983; Jusczyk, 1977; Jusczyk & Thompson, 1988; Morse, 1972). A better option is to use contrasts involving fricatives, since evidence for infants' discrimination of fricatives is mixed, to say the least. For instance, Eilers, Wilson, and Moore (1977) report that both 6- to 8-month-olds and 12- to 14-month-olds fail with /t-T/, while 2-month-olds (tested with a different method) succeeded with the same contrast, according to Levitt et al. (1988). Three-month-olds were unable to discriminate /sa-za/, but they succeeded with /as-az/ (Eilers, 1977; see also Aslin, Pisoni, & Jusczyk, 1983). While Holmberg, Morgan, & Kuhl (1977) are often cited as demonstrating that 6-month-olds can discriminate /s-S/ (e.g., Jusczyk, 1997:53; Levitt et al., 1988: 362), Nittrouer (2001) finds that few can. Using a within-subject design, Nittrouer (2001) documents that 6 out of 15 infants who demonstrated discrimination of vowel quality (either /sa-su/ or /Sa-Su/) could discriminate /sa-Sa/, while out of 8 children who could distinguish a stop voicing contrast (/ta-da/), none discriminated the sibilants.

1.3.2 Multidimensionality

It is clear that most speech sound contrasts are correlated with multiple aspects of the speech signal, many of which vary in tandem. For example, there are 16 correlates (or acoustic characteristics) that tend to coincide with the voicing contrast of intervocalic stops in English (Kingston & Diehl, 1994; Lisker, 1986), each of which could potentially be useful in making voicing distinctions (and which are integrated at some perceptual levels; Kingston, Diehl, Kirk, & Castleman, 2008). To be able to learn speech categories, then, infants must track numerous acoustic dimensions simultaneously as well as potentially integrate them. Given that the categories used in Maye et al. (2002, 2008) varied only along VOT, those results cannot answer the question of whether infants make use of frequency distributions spread along multiple dimensions simultaneously.

A great deal of research has assessed infants' performance with limited acoustic cues. Unfortunately, very little of it bears on the question of how multidimensionality affects infants' category learning. For instance, infants' discrimination abilities when presented with a subset of cues has been extensively documented. For instance, while young children are able to hear stop place of articulation distinctions based on either formant transitions (Moffitt, 1971) or burst spectrum (Miller, Morse, & Dorman, 1977), they are marginally better when presented with both burst and formant transitions cues as compared to only formant transitions (Williams & Bush, 1978; but these results were not replicated in Walley, Pisoni, & Aslin, 1984; see also Jusczyk, 1981). However, these studies may simply suggest that infants can better discriminate tokens that are acoustically more different from each other, and says little about how infants learn to attend to correlated cues.

Others have attempted to test the hypothesis that infants' perception is affected by phonetic context, in the hope of finding out whether infants compensated for subphonemic patterns due to (what may be) articulatory constraints. For instance, some investigated whether very young infants compensate for speech rate variation (e.g., Eimas & Miller, 1980a, and Jusczyk et al., 1983). Yet others documented how a preceding frication, gap, or /r-l/ affects discriminability of formant transitions (e.g. Eimas, 1985; Eimas & Miller, 1991, Levitt et al., 1988, and Fowler, Best, & McRoberts, 1990). Overall, it is not clear whether the effects reported result from infants' innate biases (see e.g., Fowler et al., 1990, and references cited), or whether they respond to more basic auditory processes (see, e.g., Lotto, Kluender, & Holt, 1997; and references therein), but it is possible that the effects found may not have been the result of learning these co-occurrences in their ambient language.

In view of the scarceness of work documenting how infants' learning is affected by the fact that speech contrasts rely on multiple dimensions, we turn to the adult literature in order to investigate whether multidimensionality may pose additional problems to the learner. The answer appears to be yes, as unidimensional categories are easier to learn than multidimensional ones (speech: Goudbeek, Cutler, & Smits, 2008; non-speech: Goudbeek & Swingley, 2007; Goudbeek, Swingley, & Smits, 2009; but see Holt & Lotto, 2006; and visual: e.g., Alfonso-Reese, Ashby, & Brainard, 2002; Ashby, Queller, & Berretty, 1999; see Lea & Wills, 2008, for arguments this "unidimensional bias" is also found in non-human animals). For example, Ashby et al. (1999) show that, in visual category learning, adults tend to rely on a single dimension, unless otherwise prompted, and Goudbeek et al. (2009) report that, in the absence of feedback, adults quickly revert to unidimensional solutions after having learned multidimensional speech categories. Since early phonetic learning is necessarily unsupervised (that is, there is no corrective feedback to use multiple dimensions), this unidimensional bias is likely to be present in infancy. If so, then the task faced in natural language acquisition is not well-represented in the stimuli used in previous research, and it becomes pressing to re-evaluate the statistical learning hypothesis with a contrast that recruits multiple dimensions.

1.4 Current research

In short, since the statistical hypothesis is theoretically attractive as a parsimonious explanation to infants' phonological acquisition, it is worthwhile to garner further empirical evidence concerning the likelihood that it may scale up to the challenges found in natural language acquisition. The present study was designed as one step in this direction, by extending previous work in two ways: using a non-salient contrast and varying multiple dimensions. The contrast between [ca] (an alveopalatal sibilant followed by a low back vowel) and [s,a] (a retroflex sibilant followed by a low back vowel) as implemented in Polish fulfills both conditions (non-salience, and multidimensionality), as explained below. It should be noted that the phonetic transcription of these Polish sounds had been subject to some debate (e.g., Maddieson & Ladefoged, 1996, described the alveopalatal as palatalized post-alveolars), which now appears to have been resolved (see Nowak, 2006; Zygis & Hamann, 2003, for discussions).

1.4.1 Salience

While the notion of "salience" appears to be intuitive, it is certainly complicated to find the reasons why certain sounds, contrast, and dimensions are more or less so (see e.g., Holt & Lotto, 2006, for a recent discussion). However, the initial purpose of this study was to test infants' category learning with sounds that were less salient than VOT and other contrasts for which very young infants (and sometimes non-human animals) exhibit categorical discrimination. The fact that discrimination reports are conflicting suggests that infants' perception of contrasts involving fricatives is generally fragile, and particularly so for sibilant place of articulation contrasts. Among the many sibilant place of articulation contrasts that exist across languages, we selected [ca] and [s,a] because cross-linguistic data suggested that these sounds, and particularly alveopalatals, could be non-salient. In the UPSID Database (Maddieson, 1984), which contains 451 geographically diverse languages, only 21 languages contain the retroflex fricative, 9 the alveopalatal fricative, and merely 2 contain both. While it is true that one cannot make too much of differences of frequency, since many factors contribute to differences in frequency (in addition to historical factors, perceptual, articulatory, and learnability factors likely play a role; see e.g., Moreton, 2008; Ohala, 2005; Schwartz, Boë, Vallée, & Abry, 1997a, 1997b), it is rather astounding that these 2 sibilants are about 10 times less frequent than, for instance, the alveolar and palatal ones (in UPSID, alveolar: 197; palatal: 189; of which 82 languages have both).

Since previous research had documented infants' failure to discriminate the alveolar-palatal place contrast, it is necessary to show that the sounds chosen are comparable to the other sibilants. In general, one expects any two sibilant places to be more similar to one another than /s-S/, given that these two are the extremes in sibilant place (Gordon et al., 2002). Nonetheless, it is worth investigating this question on the basis of existing acoustic research. [c] and [s,] tend to occur in inventories with a 3-way (but not 2- or 4-way) place contrast in sibilants, with the third segment usually being [s] (Boersma & Hamann, 2008). In this context, one may ensure that the dissimilarity involved in [c-s,] is smaller than that involved in [s-S], at least at the physical level, by assessing the acoustic distance between [s] and either [c] or [s,]. Indeed, Polish can be described as having a 3-way place contrast among voiceless sibilants, with the third place being dental, represented here as [s] for simplicity (there is an additional sibilant, a palatalized post-alveolar, that occurs in loandwords and as an allophone of the retroflex one; see Zygis & Hamann, 2003; Zygis & Padgett, 2010). Acoustic descriptions support the hypothesis that [s] is more dissimilar to [s,] than to [c], and that the acoustic distance between [s] and either [c] or [s,] is smaller than that between [c] and [s,]; in short, [s] ≠ ≠ [c] ≠ [s,] (e.g., Kudela, 1968; Jassem, 1979; Nowak, 2006; Zygis & Padgett, 2010). The reduced acoustic distance between [c] and [s,] may be particularly problematic for Polish, as in Mandarin these sounds appear to differ acoustically to a greater extent (e.g., Li, 2009; see Chiu, 2009, for a direct comparison between the two).

In brief summary, cross-linguistic frequency and acoustic analyses suggest that the Polish sounds [c] and [s,] are less salient than [s-S], which in turn have been shown to be difficult for infants to perceive. In consequence, these sounds appear as excellent choices to test the hypothesis that categories may be learned on the basis of the distributions of acoustic cues in cases of enhancement, where initial sensitivities are weak. On the other hand, it must be noted that the two sounds chosen may not, themselves, be equal in salience. In the following subsection, we investigate possible differences in perceptual salience across the two members of the contrast. However, before proceeding, we would like to repeat a very important point. It must be born in mind that, beyond possible differences across retroflex and alveopalatal, *any* sibilant place of articulation contrast chosen would fulfill the requirement of being less salient than VOT: most infants (some as old as 14 months of age) cannot reliably discriminate the even more acoustically dissimilar (and native) alveolar and palatal sibilant contrast (Nittrouer, 2001), while 1-month-olds discriminate tokens along a VOT continuum categorically (Eimas et al., 1971). Thus, regardless of possible place-specific effects, a study looking at these sibilants would make a contribution different from that of Maye et al. (2002, 2008).

1.4.2 Differences in salience between [c] and [s,]

To assess possible differences in relative perceptual salience, we follow the general method of Narayan (2008), who documented the non-salience of /ŋ/ as compared to /m/ and /n/ on the basis of (a) their typology (/ŋ/ is a great deal less common than the others), and (b) their perception; in particular, discrimination performance for pairs involving /ŋ/ was consistently lower, both in listeners whose native language has this 3-way contrast in the position tested, and others who do not.

With respect to typological differences, while both sibilant places are rather infrequent, retroflexes are less so than alveopalatals. This may not be by chance, as evidence for retroflexes being perceptually weak is unconvincing, according to Hamann (2005). For example, Hamann (2005) argues that retroflexes may be perceptually salient given that they have spread areally, similarly to clicks. We have found no report of alveopalatals spreading areally, although this is also the case for many other sounds.

Perceptual research also provides some support for a difference between these two sounds. Naturally, it would be unsurprising if [s-s,] were discriminated better than [s-c], given that the acoustic distance between the former is larger than that in the latter (at least in terms of cues in the frication portion). What is more meaningful is that alveopalatals appear to be weaker in *identification* tasks. Thus, data reported in Nowak (2006) shows that, when faced with a variety of cross-spliced stimuli, Polish speakers perform marginally worse with alveopalatals (p = .07). and significantly so when all cues are available (calculated from Table 5: Mean difference 7.7, pooled SE = 5.61; $t(7) = 2.57$; $p < .05$). This pattern of results has been replicated with Mandarin Chinese listeners (a language that also has dental, alveopalatal, and retroflex sibilants), who show lower d', higher response times, and higher biases for alveopalatals than retroflexes both for

Mandarin stimuli and for Polish stimuli (Chiu, 2010). Moreover, when confronted with a continuum between Polish retroflex and alveopalatal, Mandarin listeners tend to label a larger portion of the continuum as retroflex (McGuire, 2007; p. 74). Finally, in the same continuum identification task, American English listeners, who have no phonemic experience with [c-s,], labeled tokens near the retroflex of the continuum more consistently than those in the alveopalatal end (McGuire, 2007; this was evident in several experiments; see e.g., p. 51 and p. 73). In short, Polish [s,] appears to be perceptually more salient than Polish [c] for Polish listeners as well as for non-native listeners, both who have phonemic experience with the sounds (Mandarin) or not (English).

Given the possibility of an asymmetry between retroflex and alveopalatal sibilant learning, it was not convenient to use the design implemented in Maye et al. (2002), which assumes both categories to be equal. Maye et al. (2002) had a fixed period of exposure, and then played 2 types of trials: alternating trials, in which two tokens, one for each category, were played in succession; and non-alternating trials, in which the same tokens were presented in 2 separate trials. They then collapsed across the 2 non-alternating trials and compared infants' looking times of this average with that to the alternating trials, under the assumption that infants would simply be responding to the variability within trials.

However, if one of the categories is more salient than the other, then the non-alternating trials may not be comparable, and looking times should not be averaged across them. Indeed, Maye et al. (2008) did not repeat this procedure for learning of prevoiced-unaspirated, given that there could be differences in the perception of these two sounds. That is, in an experiment using the Conditioned Head-Turn procedure, Aslin et al. (1981) documented that the prevoiced served as a better "background" than the unaspirated one, a pattern that has since been associated with perceptual asymmetries (Polka & Bohn, 2003), such that perceptually stronger categories are worse backgrounds (similarly to what happens with the perceptual magnet effect, e.g., Kuhl et al., 1992, and discussion in Polka & Bohn, 2003: 222 and 227). Therefore, after the initial exposure period, Maye et al. (2008) used a habituation-dishabituation paradigm, in which the stronger category (the unaspirated stop) served as background or habituation stimulus, and a looking time contrast was expected for the dishabituation stimulus.

There are two important problems with this procedure which prevented us from adopting it too. First, throughout habituation infants hear additional exemplars of one of the categories, thus modifying the distribution that they are exposed to. Even though this manipulation did not affect infants' perception in Maye et al. (2008), one cannot be certain that it would have an equally null effect with non-salient categories, such as the sibilants used here. The second problem with the habituation-dishabituation test is that it assumes that discrimination of the two tokens belonging to the different categories is a sign of learning of both categories. However, a similar result could ensue if only one category is learned, as long as the other token being presented is different enough to be seen as a bad exemplar or an outlier of the learned category. For example, the asymmetry effect documented by Kuhl et al. (1992) shows that infants are unable to make discriminations within their native vowel categories to a certain extent, but are still be able to discriminate their native category from a very bad exemplar (one that is far from the prototype). In these circumstances, there is an alternative interpretation of Maye et al. (2008), according to which infants learned only *one* category, that of unaspirated stops. We do not claim this interpretation invalidates Maye and colleagues' conclusion that infants' perception was altered by the distribution of acoustic cues they heard; on the contrary, it is clear that their perception *was* affected by acoustic cue distribution. However, we do believe that the test adopted cannot document learning of *both* categories. Therefore, we opted for a design that allowed us to assess learning within each category independently. In this design, infants' looking time is measured for tokens that are closer or further away from each category of interest. In order to explain this, it is necessary to first give an account of the way the continuum was built; we therefore return to this question at the end of SS1.4.3 (see SS2.1.2 for more detailed explanations).

1.4.3 Multidimensionality

Regarding the second condition, acoustic and perceptual studies suggest that the identity of Polish retroflex and alveopalatal sibilants in syllable-initial position depends on a number of acoustic correlates, which are distributed across the frication and vocalic portions. The two sounds are reported to differ on two major acoustic properties: the distribution of energy across the frequency spectrum during the period of fricative noise; and the pattern of formant transitions (in particular the second formant, F2) immediately following the onset of voicing (e.g., Lisker, 2001). Given that both sounds occur in non-vocalic contexts (according to Zygis & Padgett, 2010), there may be sufficient cues in the frication portion for accurate identification. However, Nowak (2006) demonstrates that the acoustic characteristics of their frication portions are very similar. Moreover, Polish listeners' identification is affected by cues in the vocalic portion (Nowak, 2006; Experiment 1). Similarly, Mandarin listeners attend to both the frication and the vocalic portions in their identification (Chiu, 2010) and discrimination (McGuire, 2007) of the Polish sibilants (the same as they do for their own sibilants, Chiu, 2010).

Even though the main correlates of sibilant place identity are the centroid of the distribution of energy during the frication and onset F2 in the vowel, it is clear that other correlates may have an effect perceptually. For example, Nowak (2006) finds that Polish listeners' identification is affected not only by formant transitions early in a following vowel, but also, to some extent, by more distant cues in the vowel. In order not to make assumptions regarding which cues are perceptually relevant to infants, we dispreferred synthesizing simplified stimuli (e.g., a pole followed by synthetic F1-F3) in order to generate naturally rich stimuli, while at the same time ensuring that 2 subsets of correlates could be manipulated independently. To this end, we split the [ca] and [s,a] syllables into a frication portion and a vocalic portion, and generated one continuum for each one of those portions by mixing the signals at different levels of amplitude (the method is explained in detail in §2.1.2).

In sum, the contrast chosen allows us to test the statistical learning hypothesis with sounds whose identity depends on multiple acoustic cues spread across the frication and vocalic portions. However, there were two potential problems that must be borne in mind in order to successfully implement such a design. First, several researchers have hypothesized that there may be different biases towards certain types of cues (e.g., Nittrouer, Manning, & Meyer, 1993; Holt & Lotto, 2006, among others). This question is discussed in the following section.

Secondly, since the multiple dimensions usually covary, it is possible that infants may attend only to one. To rule out this possibility, infants in this study had access to only a subset of the cues during test, with the subset being counterbalanced across infants. In particular, half of the infants were tested on tokens differing only in the frication dimension, whereas the other half were tested on tokens differing only in the vocalic dimension. There were 3 different trial types within each of these 2 test conditions, and within each of the 2 places of articulation. These trial types were "natural" combinations of a frication and a vocalic portions both belonging to the same place of articulation; "unnatural" combinations, in which the frication and the vocalic portion belonged to different places of articulation; and "mid" combinations, where one of the portion corresponded to one place, and the other was ambiguous. Thus, for infants in the frication condition, all of the tokens in all of the Retroflex trials had unambiguous Retroflex frications, but the vocalic portions varied across the trials (they were retroflex for the natural ones, alveopalatal for the unnatural ones, and ambiguous for the mid trials). This design is further explained in §2.1.2.

Before closing this section, we would like to underline that the goal of the present research is to test the statistical learning hypothesis in a context where multiple dimensions vary. The contrast chosen fulfills this requirement, thus making a contribution to the phonetic learning literature that is different from that of previous studies. However, the studies presented here cannot contribute to many topics associated with multidimensional categories, including the existence and origin of cue-weighting biases, their change with development, the integration of correlated cues in a single percept, and whether the principles underlying unidimensional and multidimensional categories are fundamentally different.

1.4.4 Possible biases underlying cue-weighting

Even though sound contrasts are usually correlated with multiple cues, listeners typically rely primarily on a subset of those cues; in other words, listeners weight certain cues more heavily than others. Furthermore, it has been proposed that cue-weighting schemes may vary depending on several factors. Since the literature on cue-weighting is large, a few representative publications are cited for each factor: the auditory/perceptual salience of the cues themselves (Holt & Lotto, 2006; Goudbeek & Swingley, 2006; Ohde, Haley, & McMahon; Sussman, 2001); the contents of the phonemic inventory (Wagner, Ernestus, & Cutler, 2006); the phonetic context in which the contrast is presented (segmental context, Mayo & Turk, 2004, 2005; syllabic structure, Nittrouer, Miller, Crowther, & Manhart, 2000); and listeners' characteristics, (including articulatory experience, Nittrouer, 1992: 379; Nittrouer, 2002, 2006; lexical development, Nittrouer & Crowther, 1998: 268-269; literacy, Mayo, Scobbie, Hewlett, & Waters, 2003).

For the present study, given that young infants have neither a native language inventory nor a lexicon, no productive experience with sibilants (Robb & Bleile, 1994), and are preliterate, only biases based on auditory/perceptual salience and phonetic context are relevant. Thus, the question reduces to: specifically for learning the syllables [ca] and [s,a], can we predict that infants may be biased towards either vowels or frications? This question is rather difficult to answer, as there is very little work on auditory/perceptual biases affecting cue-weighting in category learning, possibly limited to Goudbeek & Swingley (2006) and Holt & Lotto (2006). Both studies report that adults learning of non-speech categories tend to weight absolute or center frequency less heavily than sweep rate (in Goudbeek & Swingley, 2006) or modulation frequency (in Holt & Lotto, 2006). This could be interpreted as a bias towards dynamic spectral characteristics, interpretation that would entail that infants should be biased towards dynamic cues, such as formant transitions. However, Bohn & Polka (2001) document that infants' discrimination of vowels when provided either only with formant transitions or only with steady centers is the same as their discrimination in the presence of both cues. We conclude, then, that there is little reason to predict that infants would be biased to either the vocalic or the frication portions in the present study.

1.4.5 Summary of the motivation and predictions

The statistical learning hypothesis predicts that infants' perception will be affected by distributions of acoustic cues in their input, and that this may form the basis of category learning. Previous research has shown that infants' perception is affected when the frequency distributions are instantiated on a single, salient acoustic dimension, VOT. The present study sought to extend that work by comparing infants' perception of two non-salient sounds across two exposure conditions, which differed in the frequency distribution of multiple acoustic cues. The stimuli chosen were the Polish syllables [ca-s,a], which fulfilled the conditions of not being salient and differing along multiple dimensions. However non-salient these sounds may be, following the statistical learning hypothesis, we predict that infants' perception will be different following the different exposure conditions. In addition, some evidence suggested that [ca] may be less salient than [s,a]. Therefore, the design was adapted to test category learning in each place separately, with the prediction that learning may be evident for the retroflex, but not the alveopalatal, place. Finally, two sets of cues were varied independently during test, those in the vocalic and the frication portions. Based on previous literature, there was little reason to expect that infants would be biased towards either the vocalic or the frication portions.

2.0 Experiment 1: Flat and Two Peak distributions

In the first experiment, two groups of infants heard one of two different distributions during an initial exposure period. In the *Flat* group, infants heard all familiarization tokens repeated the same number of times, such that no acoustic category would be promoted by the input. In the *Two peak* group, infants heard the “natural corners” of the continuum more frequently. After this initial exposure, all infants were tested using the Headturn Preference Procedure (HPP; Jusczyk & Aslin, 1995; used with 4-month-olds in Seidl & Cristiá, 2008) on their perception of tokens in different areas of the continuum. In each trial of the testing phase, infants were presented with two alternating tokens that differed on only one dimension (the other dimension was held constant). Therefore, in order to succeed, infants would have had to keep track of both dimensions simultaneously during the initial exposure, and they should be able to utilize them independently in their judgments during test.

We hypothesized that infants should show a graded pattern of looking time only for learned categories, with natural and unnatural trials being the most different, and mid trials somewhere in the middle (McMurray & Aslin, 2005). Therefore, infants should exhibit this graded looking time pattern only after the Two peak distribution. Furthermore, if infants’ perception was more easily shaped for the retroflex category, this graded pattern would be more marked for trials within the retroflex place and less so for those in the alveopalatal place. Finally, even though not predicted on the basis of the literature, infants’ perception could respond better to the vocalic or the frication dimension, and this factor was also taken into account statistically.

2.1 Methods

2.1.1 Participants

Sixty-four (32 in each condition) English monolingual, fullterm infants were included (*M* = 5.0 months, range 3.95-6.02 months, 30 female). An additional 34 infants were not included for the following reasons: failing to finish the experiment due to fussing, crying or falling asleep (16); experimenter or equipment error (6); being exposed to a language other than English (3); being premature (1); or having looking times shorter than 1 second on any given trial (8).

As noted above, using non-salient sounds and varying multiple dimensions may make learning more difficult. In order to maximize the chances of success in this unfavorable scenario, we tested 4- to 6-month-old infants, given that mounting evidence suggests that infants’ phonetic learning abilities become increasingly constrained with experience. For example, younger infants succeed at learning sound patterns that older infants do not detect (Cristiá & Seidl, 2008; Cristiá, Seidl, & Francis, in press; Cristiá, Seidl, & Gerken, in press; Gerken & Bollt, 2008; Seidl, Cristiá, Bernard, & Onishi, 2009). Infants tested here were thus younger than those in the 2 previous infant category learning studies (Maye et al., 2002: 6- and 8-month-olds; Maye et al., 2008: between 7 and 9 months).

2.1.2 Stimuli

The stimuli presented both in the initial exposure and testing were produced by modifying a pair of syllables [ca] and [s,a] produced by a male Polish speaker. The original syllables were recorded in a sound-shielded booth with a head-mounted microphone (AKG, model C420) and a Marantz PMD670 solid state recorder at 44.1 kHz sampling rate and stored in .wav format. These original syllables were selected on the basis of clarity as well as similarity to the acoustic characteristics for Polish alveopalatal and retroflex sibilants reported in Nowak (2006). The acoustic measurements for the original syllables are reported in Table 1 and a graphic depiction of the most important acoustic parameters in the fricative and vocalic portions are presented in Figures 1-2.

Table 1 *Acoustic measurements (peak in the fricative spectra, and F2 frequency at the onset and midpoint of the vowels, all in Hz) for the naturally produced syllables on which the stimuli are based.*

	[ca]	[s,a]
Fricative spectrum peak (Hz)	2890	3890
Onset F2 (Hz)	1420	1720
Midpoint F2 (Hz)	1280	1320

Insert Figure 1 about here

Insert Figure 2 about here

All modifications were performed using Praat (Version 4.5.17, Boersma & Weenik, 2005). The syllables chosen were split in two at the boundary between the fricative and the vowel as determined by the onset of the first clear glottal pulse. The two frication portions were brought to the same length by excising four 8 ms portions at 20% intervals of the total length. The vowel portions were equated in length, pitch, and RMS amplitude using Praat's manipulation object which uses the Pitch Synchronous Overlap Add (PSOLA) method to align pitch periods, first equating duration, then pitch, both to a intermediate value between the two original recordings.

Finally, the endpoint fricatives were interpolated to create a ten-step fricative continuum from one place of articulation to the other. This interpolation was done by adding up the signals at different ratios of amplitude, from a 0 retroflex - 9 alveopalatal ratio for the alveopal end, to 9 retroflex - 0 alveopalatal for the retroflex end. The same manipulation was performed on the vocalic portion. Spectrograms of the 2 endpoints for the frication and those for the vocalic portion are shown in Figure 3.

Insert Figure 3 about here

This type of interpolation was selected (rather than generating synthetic syllables) because it does not artificially reduce the complexity of the sounds in question (e.g. both the pole frequency and overall spectral shape will vary in the fricative). With respect to the vocalic portion, an important consideration is how this manipulation affects the perception of the cues in the vocalic portion. In particular, this interpolation method produces intermediate tokens that essentially have doubled formants, one from each signal. Work on formant integration, now generally known as the center-of-gravity effect (see e.g. Delattre, Liberman, Cooper, & Gerstman, 1952; Chistovich & Lublinskaya, 1979; Xu, Jacewicz, Feth, & Kristamurthy, 2004) suggests that listeners perceive a weighted average when formants are within 3-3.5 Bark. This is the case for these stimuli. Specifically, the largest difference between the two vowels was between the F2 loci which were 1420 Hz, 10.73 Bark for the retroflex and 1720 Hz, 12 Bark for the alveopalatal – a difference of 1.27 Bark. Finally, an important factor in infant studies is the naturalness of the stimuli. Given that differences in length, amplitude, and pitch had been equated prior to this manipulation, there were no artifacts, and the interpolation resulted in stimuli that sounded extremely natural to the authors. To check this intuition, the 2 endpoints of our stimuli and those of Maye et al. (2008)'s coronal series (d1-100, and t1+21) were played to 8 naive listeners of mixed language backgrounds (2 Italian, 2 Russian, 2 French, 2 English), who were asked to rate the 4 stimuli in naturalness/unnaturalness on a scale from 1 (very unnatural) to 9 (very natural). Presentation was blocked by stimulus type, and the order of presentation was counterbalanced across listeners. Average ratings for the Maye et al.'s stimuli were 7, and for our stimuli 6.6. This difference was not significant [$t(7) = .75$, $p > .48$].

The 10 fricative and 10 vocalic steps were combined orthogonally with one another yielding a bidimensional continuum of 100 tokens as represented in Figure 4. Twelve of these combination syllables were reserved for testing and the remaining 88 were presented during the initial exposure. During both phases, tokens were separated from one another by a 500 ms silence.

Insert Figure 4 about here

Therefore, infants across conditions heard the same total number of tokens (a total of 176 syllables; total presentation time was 157 seconds), and the same selection of tokens, but the frequency of specific tokens varied across the two Distribution conditions, as represented in Figure 4. Specifically, in the *Flat* distribution, infants heard every familiarization token twice, such that no combination of fricative and vowel was more frequent than other combinations. This condition represents a perceptual baseline, since infants are exposed to the same sounds but there is no mode in frequency to shape their perceptual space. In contrast, the *Two peak*s distribution suggested categories in the two natural endpoints, such that combinations of fricatives and vowels corresponding to the same category were more frequent.

Insert Figure 5 about here

2.1.3 Design and procedure

The experiment consisted of three phases, an initial exposure phase, a brief training phase, and a test phase. During the initial exposure phase, infants heard the familiarization tokens in a randomized, fixed order, while sitting on their caregiver's lap in a small room. In order to keep infants' attention and minimize distress, the auditory stimuli were synchronized with a visual display generated using the iTunes 6 viewer on a Macintosh computer projected on a large screen.

Immediately after the initial exposure, caregiver and infant went into an adjacent testing booth. This booth consisted of a 3-walled enclosure made of white pegboard panels, approximately 4.5 feet high, with white curtains that descended from the ceiling to meet the pegboard. The pegboard was backed by thick cardboard to cover the holes, except for one large and two smaller openings in the front panel. The larger opening allowed a camera to record the session. A smaller opening allowed the experimenter to view the infant's headturns. Finally, a third opening allowed a secondary observer, such as a second caregiver, to view the procedure. Both the experimenter and the caregiver holding the infant wore tight-fitting Peltor Aviation headphones through which they listened to loud masking music superimposed over low-level white noise. A chair was placed in the center of the booth, facing the front panel. A light was attached at the center of each panel, at the approximate eye level of an infant seated on a caregiver's lap in the chair. The light on the front panel was green, while the lights on the side panels were red. Directly behind each red light, there was a Cambridge Ensemble II speaker. A Macintosh G4 computer fed the audio signal through a Yamaha Natural Sound Stereo Receiver RX-49 audio amplifier to these speakers.

Testing trials began with the green light at the front flashing. When the infant oriented towards this light, it was extinguished and one of the red side lights began flashing. When the infant oriented towards the flashing side light with a 90-degree head-turn, the trial began. During any given trial, one pair of test stimuli was played through only one of the speakers, at the same time as the corresponding light was blinking. The stimuli continued to be played until the infant oriented more than 30 degrees away for longer than two consecutive seconds, or until the test tokens had been repeated 10 times. When one of these conditions was met, the trial ended, and the following trial started with the light at the front flashing. Side light and order of presentation of the stimuli were randomized by the computer program used to run the study. In this testing booth, the experimenter coded the infant's orientation towards the lights (and sound source) by means of a button box. The dependent measure was the amount of time that the infant oriented to the light in each trial (Looking Time, LT).

Upon entering this testing booth, infants first heard two *training* trials. In these first two trials, a maximum of 10 seconds of instrumental music was presented in order to introduce infants to the fact that sound presentation was contingent on their looking at the blinking light. After this brief training, infants were presented with the *test* trials proper, which reflect the combination of three variables: Dimension, Place, and Trial Type.

In order to keep the testing phase relatively short, each infant was tested on trials varying along only one *Dimension*. Half the infants were tested with trials in which tokens had the same fricative portion but the vocalic portion varied, and for the other half the vocalic portion remained constant and the fricative varied. In the first block of trials, the value of the unvarying portion was set to one place of articulation (e.g., retroflex), and in the second block of trials the value was set to the other place (e.g., alveopalatal). The order of presentation of *Place* was counterbalanced across participants.

Within each block, 4 *types* of trials were presented, all of which consisted in 2 alternating tokens. In order to refer to the tokens in each trial, syllable combinations are denoted by referring to the steps in the continuum where its parts are located. For example, the syllable f6v0 refers to the combination of the fricative portion f6, generated by adding together the frications of the retroflex and the alveopalatal tokens at a 6 to 3 ratio of amplitude, with the vocalic portion v0, generated by adding together the vocalic portions of the retroflex and the alveopalatal tokens at a 0 to 9 ratio of amplitude. In *long* trials, the tokens presented were the extremes along a single dimension and place (e.g., in the alveopalatal, fricative-varying side, f0v0 and f9v0). This trial type allowed to determine what drove infants' preference with the present stimuli and procedure. Briefly, if infants were responding to acoustic distance between the two tokens being presented in a trial, looking times to the long trials ought to be maximal (if infants exhibited a preference for distinct tokens) or minimal (if they preferred similar-sounding tokens), given that these two tokens span the largest distance, both in terms of interpolation steps and in acoustic terms.

The other three pairs spanned the same distance in terms of interpolation steps: we call these natural, mid, and unnatural. *Mid* trials consisted of the two tokens in the middle of one side (e.g., f3v0 and f6v0 for the alveopalatal, fricative side; that is, when fricative varies, and the non-varying vocalic portion cues alveopalatal place). The *natural* trials consisted of the token in a natural corner (that is, f0v0 - an alveopalatal fricative combined with an alveopalatal vocalic portion - or f9v9 - a retroflex fricative combined with a retroflex vocalic portion) and the token closest to it among the ones reserved for test, three steps away along either the vocalic or the fricative dimension. Likewise, the *unnatural* pairs consisted of a token in the unnatural corner (f0v9 or f9v0) and the test token closest to it along either dimension.

2.2 Results

Before carrying out analyses, comparison of the long trials with the other three types allowed to determine the interpretation of the looking time measure. If infants were responding to the perceptual distance between the two tokens being presented in a given trial, LT to long trials ought to be either maximal or minimal. Since this was not the case, as shown in Table 2, these trials were dropped from the analyses, as the interpretation of looking times is that of preference rather than discriminability (although a separate set of statistics showed that all factors and interactions found significant in the reported analyses remain so in analyses including these trials). Thus, infants are likely not responding to the distance spanned between the 2 different tokens within a given trial. Looking time instead depended on the general acoustic characteristics of the 2 tokens presented. For example, looking times to the natural retroflex trials are best interpreted as those responding to the natural retroflex area of acoustic space, depicted on the bottom right corner in Figure 4; looking times during the unnatural retroflex vocalic trials are due to tokens in the top right corner in the same Figure, etc.

Table 2 *Means (standard error) of looking times in seconds by Distribution, Place, and Trial Type.*

Experiment 1: Flat

Place	Long	Natural	Mid	Unnatural
Retroflex	11.3 (1.1)	13.0 (0.8)	10.4 (1.0)	11.7 (1.0)
Alveopalatal	12.3 (0.9)	10.4 (1.0)	11.9 (1.1)	10.5 (1.0)

Experiment 1: Two peaks

Place	Long	Natural	Mid	Unnatural
Retroflex	11.9 (1.1)	13.0 (1.0)	10.9 (1.1)	9.0 (1.1)
Alveopalatal	10.9 (1.1)	11.3 (1.1)	10.8 (1.1)	13.2 (1.0)

A repeated measures ANOVA with Distribution (Flat, Two Peaks) and Dimension (Frication, Vowel) as across-subject factors, and Place (Alveopalatal, Retroflex) and Trial Type (Natural, Mid, Unnatural) as within-subject factors on Looking Times as dependent measure revealed a significant three-way interaction between Distribution, Place and Trial Type [$F(2, 120) = 5.17, p = .007$], and a significant two-way interaction between Trial Type and Place [$F(2, 120) = 5.88, p = .004$], and no other effects or interactions [all F values < 2, $p > .16$]. Looking times by Place and Distribution are shown on Figure 6, to which we refer in the interpretation of the interactions. Notice that this Figure also shows the looking times in Experiment 2, which are discussed below. The three-way interaction in the present experiment arises because infants' looking times to the different trial types diverged depending on the familiarization Distribution within the Retroflex place, but not within the Alveopalatal one. This is evident by comparing the Flat and Two peaks for the Alveopalatal place (on the far left side of Figure 6), with those in the same conditions but in the Retroflex place (right-hand side of Figure 6). In other words, infants' perception may have been affected by the familiarization distribution with respect to retroflex tokens, but not alveopalatal ones. In addition, the Trial Type by Place interaction likely emerges because infants' looking times varied significantly within the Retroflex place, but not so in the Alveopalatal place. Put differently, infants may be able to hear differences across the retroflex trials, but not across the alveopalatal ones.

These interpretations were tested with a follow-up ANOVA within each Place, in order to determine differences in perception dependent on initial exposure within the same regions of acoustic space. The analysis in the alveopalatal place revealed no main effects or interactions [all *F*s < 2, except for Distribution*Type: *F*(2, 120) = 2.58, *p* > .05]. In other words, there is little evidence of shifts in perception (learning) within this region of acoustic space.

In contrast, within the retroflex place, there was a significant effect of Trial Type [*F*(2, 60) = 8.4, *p* < .001], as well as an interaction of Type*Distribution [*F*(2, 120) = 3.08, p < .05; all other *F*s < 2, *p* > .22]. Thus, infants' perception varies with the different familiarization Distributions within this region of acoustic space.

Insert Figure 6 about here

The main effect of Trial Type is due to infants showing a certain preference for natural trials across distributions. The interaction Type*Distribution was investigated through post-hoc analyses, with the alpha set at .016, for 3 comparisons within each distribution. Infants who had heard a Flat distribution appeared to display some preference for the extremes of the space, although not significantly when controlling for multiple comparisons [natural-mid: *t*(31) = 2.42, *p* > .016; unnatural-natural: *t*(31) = 1.5, *p* > .016; mid-unnatural: *t*(31) = 1.25, *p* > .016]. In contrast, after hearing a distribution with two peaks infants treat natural and unnatural trials differently [*t*(31) = 3.83, *p* < .001], although there are no significant differences between natural and mid [*t*(31) = 1.98, *p* > .016], and mid and unnatural [*t*(31) = 2.18, *p* > .016].

2.3 Discussion

The goal of this experiment was to extend previous findings regarding infants' ability to learn categories from distributions in the acoustic signal in two ways. First, two non-salient categories was chosen; and second, two dimensions were varied simultaneously during the initial exposure, although only one was informative during test, to assess whether infants were tracking both during the learning phase. With respect to multidimensionality, results suggest that it did not prove an absolute hurdle, since infants showed different looking times according to trial type in at least one condition. Furthermore, the lack of interactions with dimension is consistent with the hypothesis that infants were tracking both dimensions during the initial exposure and could rely on limited cues during testing. As for non-salience, statistical learning may extend to *some* non-salient categories. Indeed, infants' perception of natural and unnatural retroflex consonant-vowel combinations was different in the baseline as compared to the Two peak condition, suggesting some reorganization of perceptual space around the retroflex prototype. However, no learning was evident in the alveopalatal place of articulation. This difference has important consequences for our conception of how perceptual acquisition proceeds in infancy, as follows.

As mentioned in the Introduction, perceptual acquisition of non-salient sounds may involve *enhancement*, where initially weak abilities are tuned, or *induction*, where there is no evidence of initial abilities. Based on the evidence summarized on SS1.4.1-2, while both sibilants used in the present study are non-salient, some evidence suggests that retroflexes are not inherently weak. Additionally, looking times to natural retroflex combinations were longer than less natural combinations after the simple exposure to the Flat distribution. This may indicate that infants are sensitive to the well-formedness of this combination even in the absence of experience with the distributions encountered in natural languages; more importantly, this indicates that infants start out with a fine-grained perceptual sensitivity to distinctions within those regions of acoustic space. In the Two peaks condition, repeated presentation of tokens near natural retroflex combinations has emphasized the pre-existing (dis)preferences, with long looking times towards natural retroflex combinations and shorter ones to the unnatural retroflex combinations. Thus, the case of retroflex appears to be an example for *enhancement*: initially weak sensitivities are ameliorated by exposure to statistical distributions.

Alveopalatals present a different case, since no such baseline preference for the alveopalatal natural combinations were evident after the Flat distribution. Interestingly, no learning appears to have taken place in this area of perceptual space after the Two peaks distribution. In this context, one can liken the situation to that of induction, that is, perceptual learning in the absence of prior sensitivities. However, there is an alternative explanation that ought to be ruled out before entertaining this possibility. One confounding factor in the present experiment was that infants not only had to learn a non-salient category in the absence of obvious pre-existing abilities; but also initial exposure included a competing retroflex category, for which infants appear to have some predilection. More in general, if 2 categories occur in the input, in the presence of limited resources, only the more salient one may be learned.

As mentioned in SS1.4.2, there is some evidence suggesting that the retroflex sibilant tends to dominate perceptual judgements, since Polish and Mandarin listeners' identification of alveopalatals and retroflexes is asymmetrical. Furthermore, the potential dominance of retroflex has been documented with the stimuli used in the present experiment, which are drawn from McGuire (2007). There, it was reported that listeners whose language background contains both alveopalatal and retroflex sibilants tested on the same stimuli used in Experiment 1 tend to allocate a more restricted area in acoustic space to the alveopalatal category (see, e.g., the Polish and Mandarin listeners in the left and right panels of Figure 7). Even listeners whose ambient language do not contain these two categories tend to label the retroflex end of the continuum more consistently than the alveopalatal one (see the English labelers in the middle panel of Figure 7).

---

Insert Figure 7 about here

---

Thus, it is still possible that infants could learn the alveopalatal category in the present setup provided that there is no competition from a similar category. A second experiment was carried out to assess this possibility, by presenting infants with numerous tokens around the alveopalatal natural categories only. We predicted that, if infants succeeded in learning this contrast, a significant effect of Trial Type should be found, such that, similarly to that encountered for the retroflexes, infants would exhibit a preference for the natural combinations of alveopalatal frication and vocalic portions.

3.0 Experiment 2: Alveopalatal category

In this experiment, infants heard tokens near the alveopalatal natural combinations much more frequently than any other combination. While no changes could be expected in the retroflex end of the continuum, the design was the same as that in Experiment 1, in order to allow for a comparison of performance with infants in the Flat condition of Experiment 1.

3.1 Methods

3.1.1 Participants

Thirty-two English monolingual, fullterm infants were included (M = 4.99 months, range 4.18-5.82 months, 21 female). An additional 8 infants were not included for the following reasons: being exposed to a language other than English (7); or having looking times shorter than 1 second on any given trial (1).

3.1.2 Stimuli

The same stimuli were used as in Experiment 1.

3.1.3 Design and procedure

The same design and procedure was used as in Experiment 1, except that the frequency of specific tokens during familiarization was altered, as represented in Figure 8. All tokens were presented only once, except those closest to the natural combinations, which were presented more frequently.

Insert Figure 8 about here

3.2 Results and discussion

A repeated measures ANOVA within the alveopalatal place with Dimension (Frication, Vowel) and Type (natural, mid, unnatural) revealed no main effects or interactions [all *F*s < 1]. Furthermore, an ANOVA including the Flat condition revealed no effect of Distribution nor interactions with it, suggesting that performance in the alveopalatal place was not statistically different after a Flat familiarization and after a familiarization with a highly kurtotic unimodal distribution centered over the alveopalatal end of the continuum. Looking times for this experiment are reported on Table 3.

Table 2 *Means (standard error) of looking times in seconds by Place and Trial Type. Looking times for the retroflex place are reported here for completeness. There are no differences from the Flat condition in Experiment 1 in either place.*

	Experiment 2: Alveopalatal		
Place	Natural	Mid	Unnatural
Alveopalatal	11.3 (1.1)	10.8 (1.1)	13.2 (1.0)
Retroflex	12.18 (1)	9.36 (.9)	11.76 (.9)

Thus, it was not the presence of the competing category (the retroflex category) that prevented infants' learning of the alveopalatals in the Two Peaks condition. Infants failed to show any significant preference within the block where the constant dimension was consistent with an alveopalatal place of articulation, even after hearing many more repetitions of alveopalatal tokens. Notice, for example, that the proportion of repetitions used was much higher than in the Two Peaks condition (which is 13 times as many as that of infrequent tokens) and than in previous work (e.g., 4 times as many as that of infrequent tokens in Maye et al., 2002, and Maye et al., 2008). Nonetheless, it is still possible that our training simply was not long or intense enough. In order to prove this explanation, it would be necessary to carry out additional experiments increasing the length of exposure until a non-null result was found. Since the process of lengthening the exposure within an experimental setting could continue *ad infinitum,* we leave that endeavor for future research, and advance the provisional conclusion that some non-salient categories are more amiable to learning than others. We return to this question below.

4.0 General discussion

Previous research (Maye et al., 2002; Maye et al., 2008) has provided convincing evidence that infants' sensitivities can be shaped even by brief exposure to the distribution of a single, salient acoustic cue in a simplified perceptual space. The goal of the present study was to assess the effects of exposure to different distributions of acoustic cues on infants' perception of a pair of *non-salient* sibilants, which were cued by *multiple dimensions*. Results suggest that multidimensionality did not pose a problem, while different learning outcomes ensued for the two sibilants. Each of these findings are discussed in more detail in the next 2 subsections, and their theoretical implications are drawn out in SS4.3.

4.1 Effects of salience

In the present study, infants' perception of 2 sibilants differing in place of articulation was assessed after different exposures to a continuum between them. In a condition that acts as baseline, infants were simply exposed to the whole continuum. After this exposure, they exhibited somewhat longer looking times to natural retroflex combinations, while no such preference appeared in the alveopalatal trials. In a second condition, infants heard many more tokens in the acoustic area corresponding to natural retroflex and natural alveopalatal tokens. After this experience, infants' preference for natural retroflexes and dispreference for unnatural retroflex combinations reached statistical significance, a result compatible with a reorganization of perceptual space triggered by learning through exposure to acoustic cue distributions. In stark contrast, no such learning occurred for the alveopalatal series. In view of the repeatedly documented fact that sibilants are challenging for infants (see SS 1.3.1), the retroflex results support the hypothesis that infants can learn some non-salient categories by relying on frequency distributions in acoustic space, while the difference between retroflexes and alveopalatals in the baseline and the experimental conditions may shed light on additional effects of salience.

Indeed, pre-existing sensitivities involving retroflex tokens may have enabled infants to attend to the frequency distributions in this region of acoustic space, thus constituting a necessary condition for learning. That is, infants' attention to natural retroflex combinations could have acted as an anchor for the distributions encountered in the input. Given the lack of preferences for the alveopalatal tokens in the baseline condition, no such perceptual bootstrapping could happen for the alveopalatal sibilants. It may not be by chance, then, that there was no evidence of learning after exposure to alveopalatal tokens. We return to ways in which this situation may be resolved in SS4.3.

In short, the present study extends the effects of exposure to acoustic cue distributions beyond the realm of VOT. These results strengthen the power of the statistical learning explanation to encompass dimensions that have not been documented as being psychoacoustically salient. At the same time, they do not rule out the possibility that statistical distributions are not a sufficient condition for learning to take place, but that a sensitivity to the acoustic dimensions involved may be a necessary condition instead.

4.2 Effects of multidimensionality

While a great deal of work has investigated the effect of multidimensionality on adult perceptual learning, the infant literature has lagged behind, with only a few studies documenting infants' discrimination abilities in the presence of limited cues. In this context, the present study provides a very first insight, by testing infants' ability to learn categories based on multiple varying acoustic dimensions.

Specifically, the alveopalatal and retroflex categories were cued through varying frication and vocalic dimensions during the initial exposure. Naturally, in a multidimensional continuum where dimensions are well correlated, as in the present case, listeners are still free to attend to only one dimension. Therefore, in order to be able to determine whether infants do attend to both dimensions, only one of them was available during testing, and the other rendered was uninformative by keeping it constant throughout the testing block. Given that infants' looking times to the test trials varied depending on the initial exposure they heard, this suggests that their perceptual adjustments took into account the distributions along both dimensions. Thus, our study suggests that infants *do* track acoustic distributions along multiple dimensions.

These results constitute a very first step in approaching multidimensional category learning in infancy, and they necessarily leave open a myriad questions that have occupied the field of adult category learning. It may be useful to point out three specific questions related to multidimensional speech categories that cannot be answered within the present study. First, we mentioned that there were no interactions with the factor Dimension, suggesting that performance was not markedly worse when either the vocalic or the frication dimension was available. Naturally, this may have been due to ceiling or floor effects, and it does not preclude that in a different testing situation there may be a difference between infants' ability to rely on the 2 types of information. In other words, the absence of a difference found here should not be interpreted as evidence of absence of a *weighting* bias. Secondly, it also remains for future research to determine whether infants, like adults, perceptually integrate co-occurring cue values to the point that discrimination of tokens containing *conflicting* cues is markedly worse than stimuli containing *correlated cues*. Finally, as summarized above, multidimensional categories appear to be harder for adults than unidimensional ones in speech. Furthermore, this is also the case for adult learning of non-speech and non-auditory categories, and other work suggests a similar bias in non-human animals, all evidence converging towards *unidimensional categorization* being the default. Given that we did not compare unidimensional and multidimensional categories, our data cannot contribute to this question directly. However, there is some evidence that multidimensionality, or possibly multimodality, actually boosts infants' learning. We return to this possibility in the next subsection.

4.3 Implications for theories of infant phonetic acquisition

As explained in the introduction, the statistical learning hypothesis may provide a parsimonious and comprehensive explanation for the multiple processes involved in the developmental changes in phonetic perception documented in the first year of life. Previous laboratory training studies have begun to provide empirical support to this hypothesis, by showing that infants' perception of VOT is shaped by the distributional properties of the input they are exposed to. In particular, Maye and colleagues have documented 2 of the 4 postulated processes: maintenance (Maye et al., 2002, 2008, bimodal condition), and attenuation (Maye et al., 2002, 2008, unimodal conditions), by testing category learning along a dimension that is psychoacoustically salient.

An equally important test of the statistical learning hypothesis is whether it can accommodate learning of initially weak, but nonetheless existent, sensitivities. The case of retroflexes in the present paper provides this evidence, extending previous findings to a third learning path, that of enhancement.

In contrast, the last type of statistical learning remains elusive: does pure induction based on acoustic cue distributions ever take place? In this paper, we found that for one non-salient category, there was no evidence of prior sensitivity, and no evidence of statistical learning. One may argue that the developmental pattern documented for /n/ and /ŋ/ provides evidence for acoustically based induction, as younger infants fail entirely to make this distinction, while 12-month-olds succeed (Narayan et al., 2009). However, while at this age infants probably cannot make use of top-down lexical knowledge, they do have access to information other than acoustic cue distributions, as the acoustic signal can be yoked to visual information for visible articulators, and to proprioceptive information for the sounds that infants normally babble. Indeed, visual cues are available for the /n-ŋ/ distinction (Grant & Walden, 1996; Johnson, DiCanio, MacKenzie, ??), and infants are reported to babble velar nasal consonants (Stoel-Gammon, 1985??), such that both types of information may underline the acoustic cue distributions associated with these categories.

Instead, more appropriate evidence would come from cases like /d/ and /ð/ (Polka et al., 2001), and /s/ and /S/ (Nittrouer, 2001), where neither visual nor proprioceptive information are available to yoke infants' experience with acoustic cue distributions. Interestingly, neither of these cases is resolved by 12 months, lending indirect support for the possibility that at least some non-salient categories require the conjunction of multiple types of information in order to be learned. The possibility that infants should rely on multiple types of information to learn acoustically fragile categories is not in opposition to the idea that multidimensional categories are harder than unidimensional ones, which appears to be the case for adults (and possibly non-human animals; see SS1.3.2). On the contrary, there is an important difference between multidimensionality in a single modality, and having access to multiple correlated cues in different modalities, which can contribute to heighten attention and indirectly improve performance (Bahrick & Lickliter, 2000). Previous research suggests that infants benefit from intersensory redundancy when learning perceptual rhythmic categories (Gogate & Bahrick, 1998), abstract patterns (Frank, Slemmer, Marcus, & Johnson, 2009), and sequential order (e.g., Lewkowicz, 2004), finding it easier to learn, generalize, and discriminate when multiple senses are involved.

5.0 Conclusions

In short, previous evidence supports the hypothesis that maintenance and decline of perceptual sensitivities may all be driven by frequency distributions of acoustic cues. The present study extended the power of statistical learning on the basis of distributions of acoustic cues to cases of *enhancement* and to a multidimensional context. Results further suggest that infants' prior sensitivity to the acoustic dimensions involved may be a prerequisite for learning to occur. Thus, it is unclear whether an acoustically-based statistical learning mechanism can lead to induction, since induction on the basis of acoustic information only has not been demonstrated neither in the present study nor in reported developmental changes in infant speech perception. Future work may shed light on whether infants can combine information from other senses to resolve phonetic categorization of fragile acoustical categories. In the meanwhile, we are left to conclude that there is little evidence for pure induction of categories based on frequency distributions in perceptual space, but that, otherwise, infants are still remarkable statisticians.

References

Aslin, R. N., & Newport, E. (2009). What statistical learning can and can't tell us about language acquisition. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language* (p. 15-30). Hove, UK: Psychology Press.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2: Perception, p. 67-96). New York: Academic.

Boersma, P., & Weenik, D. (2005). *Praat: Doing phonetics by computer* [Computer program]. (Retrieved May 26, 2005, from http://www.praat.org/)

Bohn, O. S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America, 110*, 505-515.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*, 207-239.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*, 159-199.

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*, B69-B77.

Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel color. *Word 8*, 195–210.

Chistovich, L. A., & Lublinskaja, V. V. (1979). The center of gravity effect in vowel spectra and critical distance between the formants. *Hearing Research, 1*, 185-195.

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61*, 1321-1336.

Eimas, P., Siqueland, E., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics, 62*, 1668-1680.

Hardison, D. (2003). Acquisition of second language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*, 495-522.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Infant discrimination of two- and five-formant voiced stop consonants differing in place of articulation. *Journal of the Acoustical Society of America, 62*, S99.

Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics, 36*, 28-54.

Kuhl, P. K., & Miller, J. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63*, 905-917.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 361-368.

Lisker, L. I. (1986). 'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*, 3-11.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*, 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can effect phonetic discrimination. *Cognition, 82*, B101-B111.

Mayo, C. (2000). *The relationship between phonemic awareness and cue weighting in speech perception: Longitudinal and cross-sectional child studies.* Unpublished doctoral dissertation, Queen Margaret University College.

Mayo, C., Scobbie, J. M., Hewlett, N., & Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *Journal of Speech, Language, and Hearing Research, 46*, 1184-1196.

Mayo, C., & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America, 115*, 3184-3194.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *Journal of the Acoustical Society of America, 118*, 1730-1741.

McGuire, G. (2007). *Phonetic category learning*. Unpublished doctoral dissertation, Ohio State University.

Mills, A. (1987). The language of blind children: normal or abnormal? *First Language, 7*, 242.

Narayan, C. (2006). *Acoustic-perceptual salience and developmental speech perception*. Unpublished doctoral dissertation, University of Michigan.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 112*, 711-719.

Nittrouer, S., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues change with linguistic experience. *Journal of the Acoustical Society of America, 94*, S.1865.2.

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America, 101*, 2253-2266.

Nittrouer, S., Miller, M. E., Crowther, C. S., & Manhart, M. J. (2000). The effect of segmental order on fricative labeling by children and adults. *Perceptual Psychophysics, 62*, 266-284.

Nowak, P. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics, 34*, 139-152.

Ohala, J. J. (2005). Phonetic explanations for sound patterns. Implications for grammars of competence. In W. J. Hardcastle & J. M. Beck (eds.) *A figure of speech. A festschrift for John Laver*. London: Erlbaum. 23-38.

Padgett, J., & Zygis, M. (2003). The evolution of sibilants in Polish and Russian. *ZAS Papers in Linguistics, 32*, 155-174.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology.

Language and Speech, 3, 115-154.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and non-categorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55*, 328-333.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109*, 2190-2201.

Polka, L., & Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20(2), 421-435.

Robb, M. P., & Bleile, K. M. (1994). Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics and Phonetics, 8*, 295-320.

Saffran, J. (2009). Acquiring grammatical patterns. In J. Colombo, P. McCardle, & L. Freund (Eds.), Infant pathways to language (p. 31-48). Hove, UK: Psychology Press.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophones and phonemes in infants' learning of sound patterns. *Language Learning and Development*, *5*(3), 191-202.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 73*, 1011-1022.

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

Zygis, M., & Hamann, S. (2003). Perceptual and acoustic cues of polish coronal fricatives. *Proceedings of the XVth International Congress of Phonetic Sciences*, 395-398.

Zygis, M. & Padgett, J. (2010). A perceptual study of Polish fricatives, and its relation to historical sound change. *Journal of Phonetics, in press, corrected proof.*

Figure Captions

Figure 1

Graphic representation of the most relevant acoustic cue in the fricative continuum, the noise spectrum, as rendered with cepstral smoothing with a 500 Hz bandwidth. The darkest line represents the alveopalatal end, the dotted line the retroflex end, and the light grey ones the intermediate tokens f3 and f6.

Figure 2

The most relevant acoustic cue in the vocalic continuum is the second formant. Since the intermediate steps (e.g., v3, v6) essentially have two formants, and following the hypothesis that listeners perceive a weighted average of them based on their amplitude (known as the center of gravity effect; Chistovich & Lublinskaja, 1979), represented here are the estimated perceptual second formant measured at 25, 50, 75, and 100 ms into the vowels. The darkest line represents the alveopalatal end, the dotted line the retroflex end, and the light grey ones the intermediate tokens v3 and v6.

Figure 3

Spectrograms of the endpoint frications (on the left panels) and vocalic portions (on the right panel). The top graphs show the alveopalatal tokens (step 0 of the continuum) and the bottom ones the retroflex ones (step 9)

Figure 4

Stimuli design: Each circle represents a syllable. Each syllable is the result of the combination of one fricative portion and one vocalic portion taken from the continua. Circles with darker outlines were not presented during the initial exposure, but instead reserved for test.

Figure 5

Frequency with which each token was presented during the initial exposures of Experiment 1, in the ‘flat' distribution condition (left), and the ‘two peaks' distribution condition (right).

Figure 6

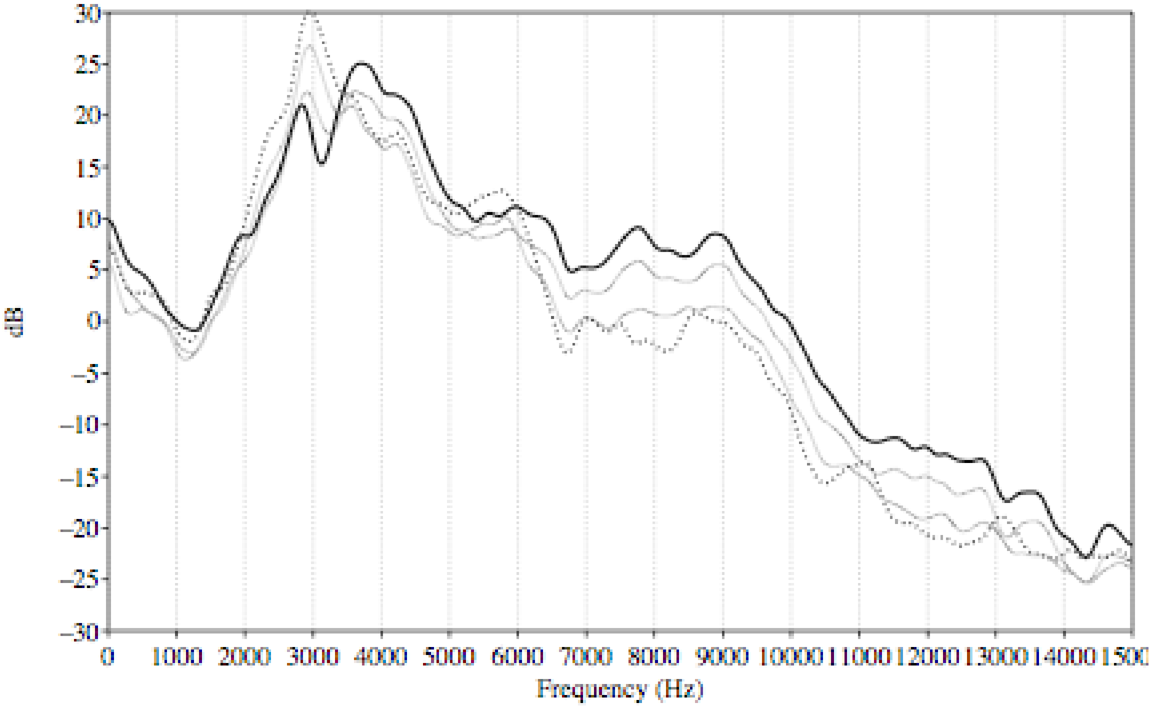
Looking times within the alveopalatal region after familiarization with one of the three distributions used across Experiments 1 and 2, and looking times within the retroflex region after familiarization with the two distribution conditions in Experiment 1. Error bars indicate standard error.

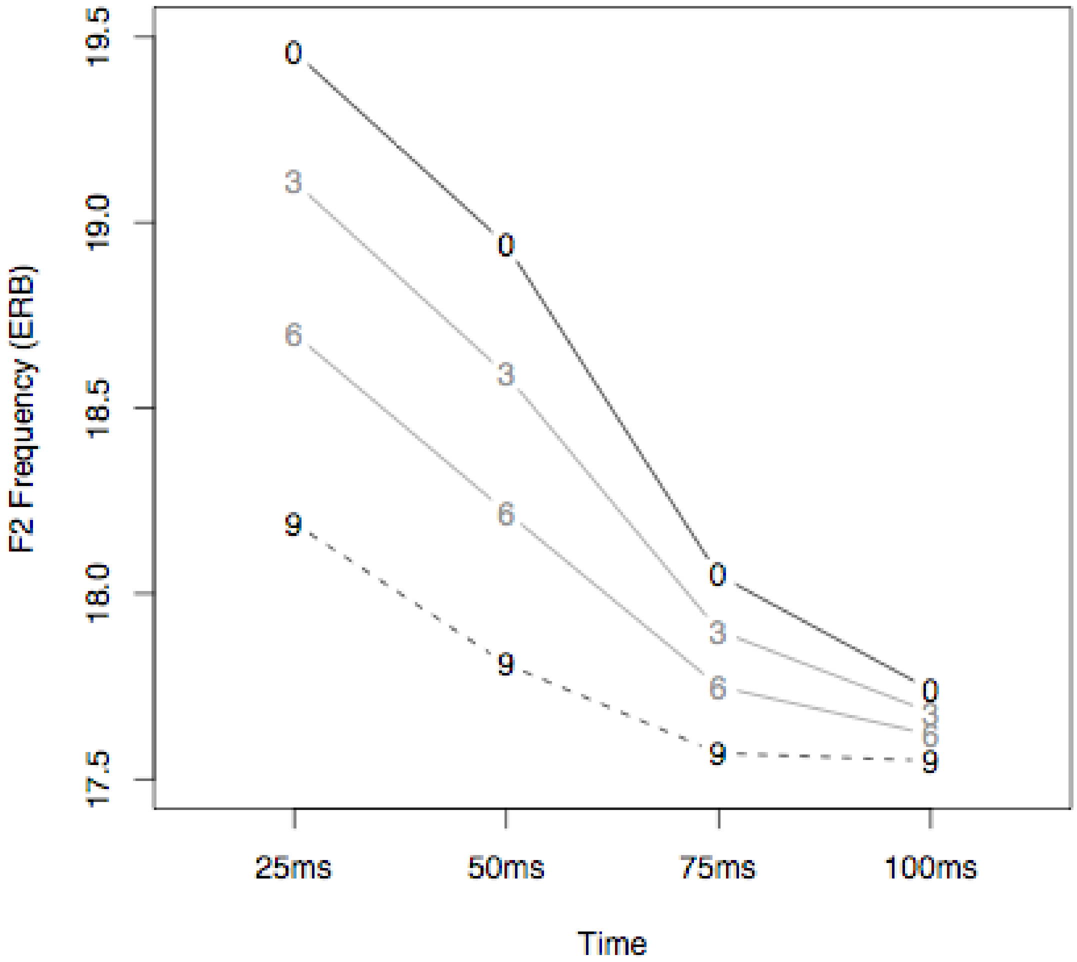
Figure 7

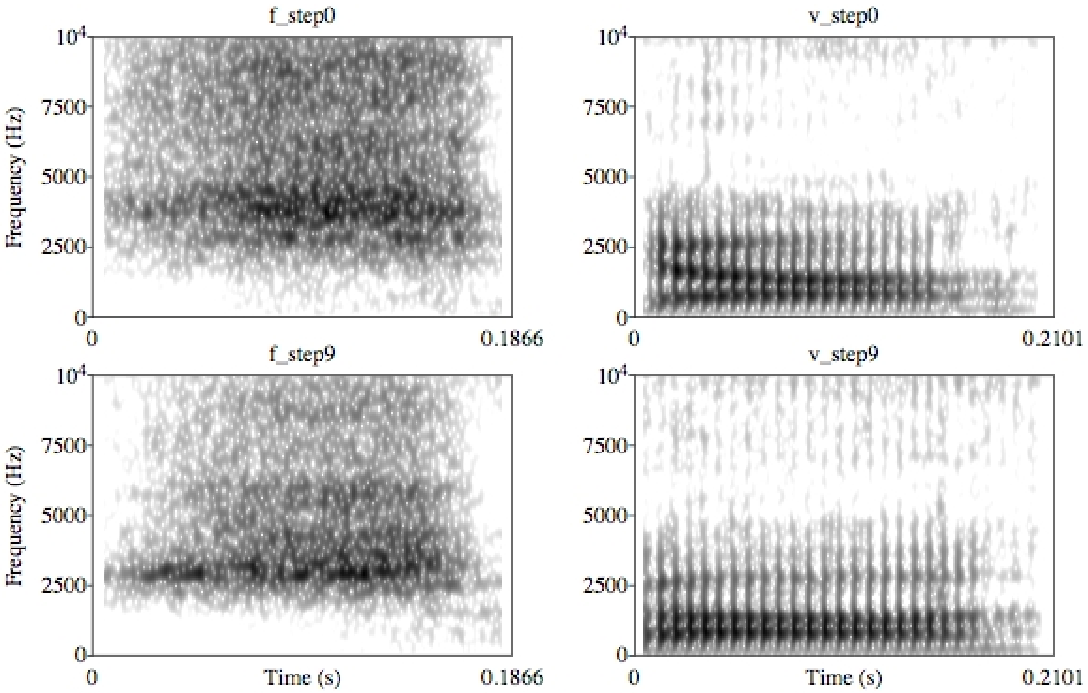
Darker squares indicate that more than two thirds of the time a stimulus was labeled alveopalatal, white squares indicate that over two thirds of the time it was labeled retroflex, and grey squares indicate that it received either label between one and two thirds of the time. The bidimensional continuum is labeled as representing two different categories by all listeners, but the specific location of the boundaries changes across language groups. Reproduced from McGuire (2007).

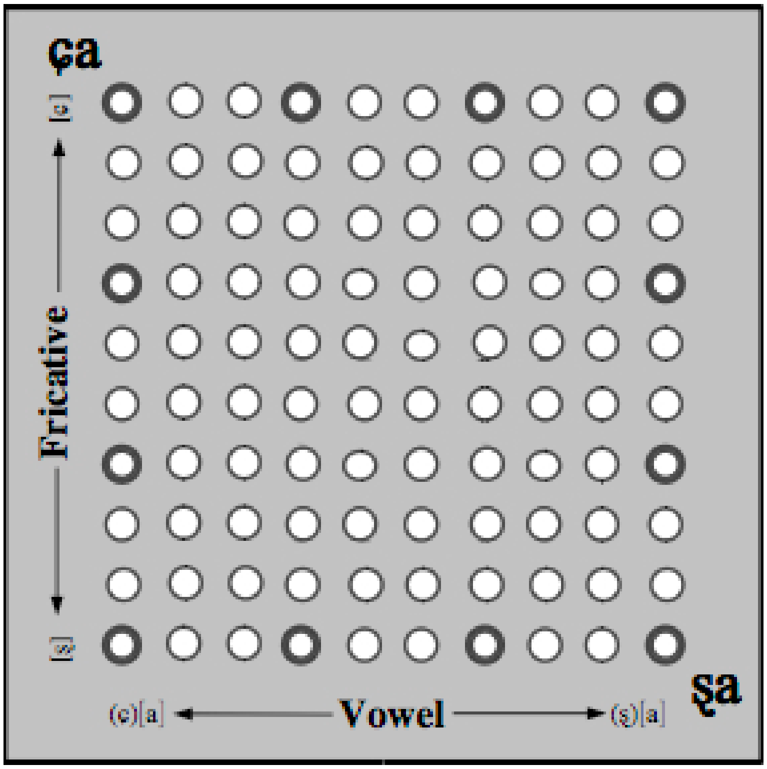
Figure 8

Frequency with which each token was presented during the initial exposure of Experiment 2, which cues the alveopalatal category.

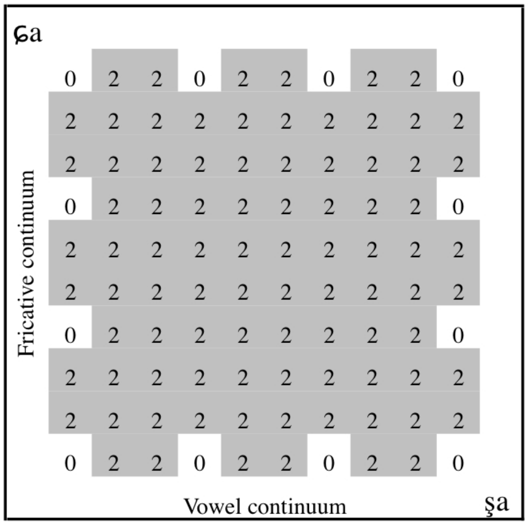

30
25
20
15
10
5
0
−5
−10
−15
−20
−25
−30
dB
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
1500
Frequency (Hz)


F2 Frequency (ERB)
19.5
19.0
18.5
18.0
17.5
Time
25ms
50ms
75ms
100ms
0
3
6
9


f_step0
v_step0
f_step9
v_step9
Frequency (Hz)
Time (s)
10⁴
7500
5000
2500
0
0.1866
0.2101

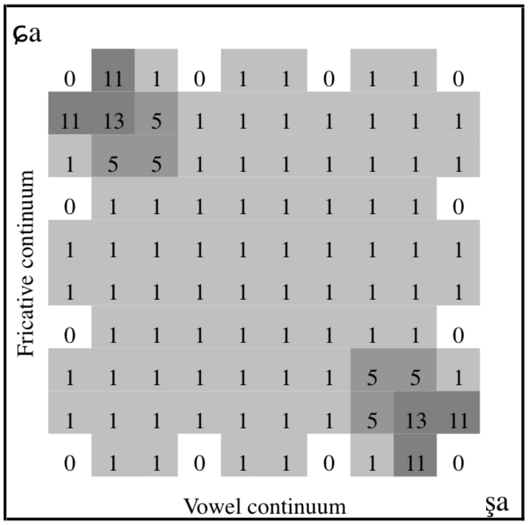

ça
[ç]
Fricative
[ş]
şa
(ç)[a]
Vowel
(ş)[a]

Flat Distribution

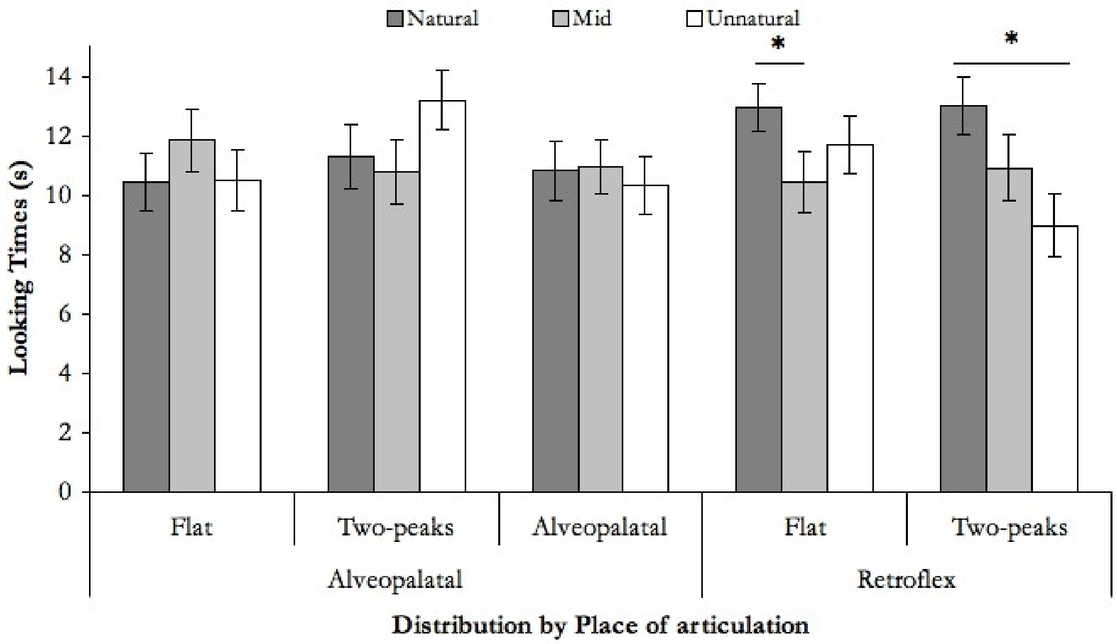


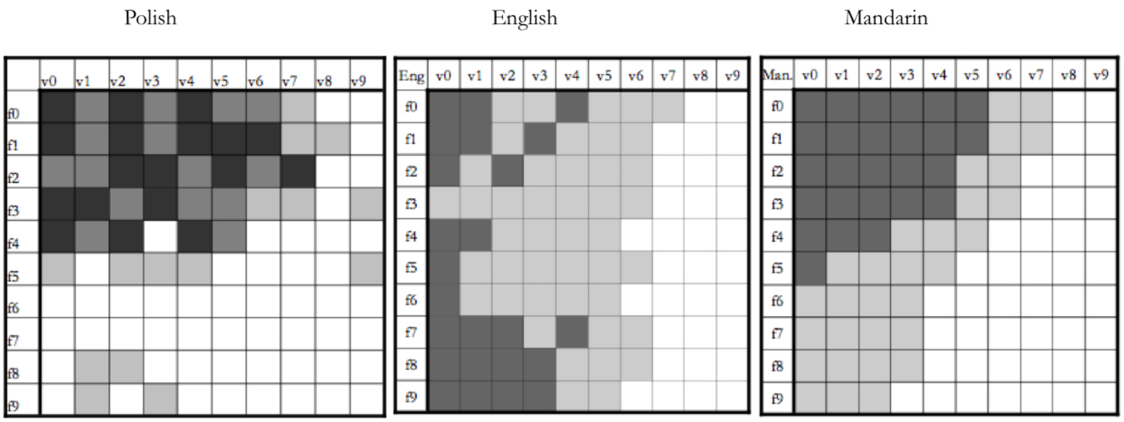
0	2	2	0	2	2	0	2	2	0
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
0	2	2	2	2	2	2	2	2	0
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
0	2	2	2	2	2	2	2	2	0
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2
0	2	2	0	2	2	0	2	2	0

Two Peaks Distribution



0	11	1	0	1	1	0	1	1	0
11	13	5	1	1	1	1	1	1	1
1	5	5	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	5	5	1
1	1	1	1	1	1	1	5	13	11
0	1	1	0	1	1	0	1	11	0


Natural
Mid
Unnatural
*
*
Looking Times (s)
0
2
4
6
8
10
12
14
Flat
Two-peaks
Alveopalatal
Flat
Two-peaks
Alveopalatal
Retroflex
Distribution by Place of articulation

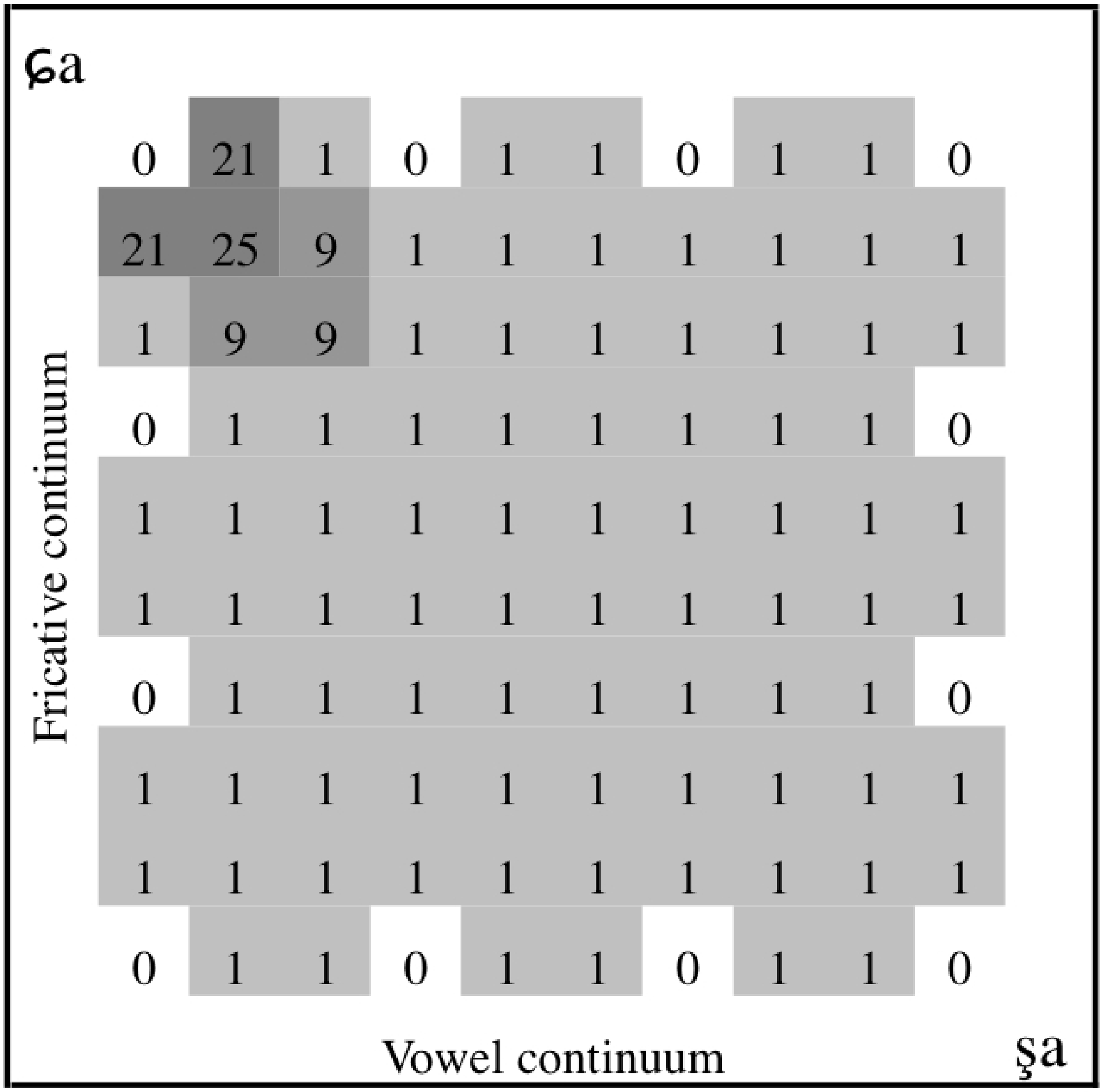


Polish

English

Mandarin

Experiment 2



0	21	1	0	1	1	0	1	1	0
21	25	9	1	1	1	1	1	1	1
1	9	9	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
0	1	1	0	1	1	0	1	1	0