To appear in the Proceedings of the 27th Annual Boston University Conference on Language Development. Sommerville, MA : Cascadilla Press.

Allophonic Variation and the Acquisition of Phoneme Categories

Sharon Peperkamp[12], Michèle Pettinato[1], and Emmanuel Dupoux[1]
[1]Laboratoire de Sciences Cognitives et Psycholinguistique, Paris
[2]Université de Paris 8

1. Introduction

One of the first tasks infants have to accomplish in order to acquire their native language consists of segmenting the continuous signal into discrete categories that represent the vowels and consonants used in their language. Experimental research has shown that this segmentation is acquired during the first year of life. In particular, vowel categories are established at around 6 months (Kuhl *et al*. 1992; Polka & Werker 1994), and consonant categories at around 10-12 months (Werker & Tees 1984a). Recently, it was argued that a statistical analysis of the acoustic space and the formation of prototypes might be used by infants to acquire these categories (Kuhl *et al*. 1997; Maye, Werker & Gerken 2002).

Within a language's segment inventory, a distinction is made between *phonemes*, i.e. those segments that are used to make lexical distinctions, and *allophones*, i.e. phonetic variants of phonemes that appear in certain phonological contexts. Consider, for instance, the following segments occurring in English: [t, d, t^h]. The distinction between [t] and [d] is phonemic (witness the minimal pair *hat* - *had*), whereas [t^h] is an allophone of the phoneme /t/ that occurs at the beginning of stressed syllables (Kahn 1976). The distinction between phonemes and allophones is to a large extent arbitrary. For instance, Korean has a phonemic distinction between [t] and [t^h] and an allophonic one between [t] and [d] (Kim 1990).

To the extent that allophones are realized as distinct acoustic prototypes, a statistical algorithm like the one proposed by Maye, Werker & Gerken (2002) yields the acquisition of segmental categories, not of more abstract phoneme categories. Thus, it predicts that both English and Korean infants acquire three separate categories for [t], [d], and [t^h]. Several articles, though, have shown that adults do not process allophonic contrasts in the same way as phonemic contrasts, suggesting that at one point they have acquired the distinction between

* Research for this paper was supported by grants from the CNRS ('Aide à Projet Nouveau' and 'Cognition et Traitement de l'Information'), the French Ministry of Research ('ACI Cognitique 1999'), the Swiss Academy for Humanities and Social Sciences ('Young Researcher Grant'), and the Advanced Telecommunication Research Institute Japan. We would like to thank Jessica Maye for comments and discussion.

phonemes and allophones. For instance, Pegg & Werker (1997), using an AX discrimination task, showed that adults' performance on the allophonic contrast between voiced [d] (as in *day*) and the voiceless unaspirated [t] that occurs after [s] only (as in *stay*), while better than chance, is worse than that on a phonemic contrast. Likewise, Whalen, Best & Irwin (1997) used an AXB discrimination paradigm to study perception of the allophonic contrast [p-pʰ] (with [p] as in *happy*), compared to that of the phonemic contrasts [b-p] and [b-pʰ]. They found that perception of the allophonic contrast was worse than that of the phonemic ones. Finally, in a gating experiment, Lahiri & Marslen-Wilson (1991) showed that adult native speakers of English, but not of Bengali, interpret nasality on vowels as an indication that a nasal consonant follows. Indeed, in English nasal vowels only occur as allophones of oral vowels before nasal consonants, whereas in Bengali the distinction between oral and nasal vowels is phonemic.

Given these findings, it is important to examine how infants might acquire the distinction between phonemes and allophones, and hence reduce the inventory of segmental categories to an inventory of abstract phoneme categories. In principle, infants might rely either on a minimal pair analysis or on a distributional analysis. The former would involve establishing a list of words with different meanings; those pairs of words that differ only in one segment contain a phonemic opposition. Hence, the presence of a substantial lexicon would be a prerequisite for this strategy to work. By contrast, a distributional analysis would exploit the fact that phonemes and their allophones have complementary distributions. This strategy would involve establishing for each segment a list of phonological contexts in which it appears. Pairs of segments whose lists of contexts have an empty intersection have complementary distributions and hence involve an allophonic distinction. This strategy, then, could be applied at an earlier age than the minimal pair analysis, i.e. as soon as segment categories have been established.

There is no experimental evidence concerning the age at which the distinction between phonemes and allophones is acquired. However, we previously argued that learning phonemes and allophones on the basis of a distributional analysis is worth exploring, since it has the advantage that it makes lexical acquisition easier (Peperkamp & Dupoux 2002). That is, words can surface in more than one phonetic form, due to allophones that appear in word-initial or word-final position and that are conditioned by the presence of certain segments in the preceding or following word, respectively. For instance, the phoneme /t/ is realized as [t] in *white* but as the allophone [ɾ] in *white owl*, due to the fact that it is intervocalic; knowing that [waɪt] and [waɪɾ] are realizations of a single word would lead infants to construct a single lexical entry for these forms instead of two, and hence to look for a single word meaning.

We report two experiments concerning the perception and the acquisition of phonemic versus allophonic distinctions, respectively. In the first experiment, we examine whether adult listeners exhibit differences in the perception of these two types of contrast. In the second experiment, we explore whether adult listeners can rely upon distributional information in order to construct abstract phoneme categories.

2. Experiment 1: Perception of Phonemic versus Allophonic Contrasts

Our experimental contrast consists of a uvular voiced fricative [ʁ] versus a uvular voiceless fricative [χ]. In French, the latter is an allophonic variant of the former: it is only found adjacent to voiceless consonants (cf. *poutre* [putχ] 'beacon' - *poudre* [pudʁ] 'powder', and *perte* [pɛχt] 'loss' – *perde* [pɛʁd] 's/he loses$_{SUBJ}$'). Hence the uvular voiced fricative assimilates in voicing to a preceding or a following consonant.[1]

In our first experiment, we address the following issues. Firstly, how well is the allophonic contrast perceived? Does this vary according to whether it is presented in isolation or in a phonological context? Secondly, is perception of the allophonic contrast affected depending on whether it is presented in a phonotactically legal string (following the assimilation rule) or an illegal one (violating the assimilation rule)? To examine these questions we assess the perception of the allophonic contrast and compare it to the perception of a contrast that is phonemic in French, i.e. [m]-[n], both in isolation and in context.

2.1 Method
2.2.1 Stimuli

The items for the allophonic contrast were [aʁ], [aχ], [iʁ], [iχ], [uʁ] and [uχ], and those for the phonemic contrast [am], [an], [im], [in], [um] and [un]. The phonological context was provided by 40 CV syllables consisting of one of the obstruents [p,b,t,d,f,v,s,z] followed by one of the vowels [a,i,u,e,o].

The stimuli were recorded by a male French native speaker and digitized on a PC computer at 16 kHz and 16 bits. The stimuli were spoken as [VC.CV] phrases, the first syllable being one of the test syllables and the second one a context syllable. To minimize co-articulation, a pause of about 300ms was made between the first and the second syllable. The two adjacent consonants always agreed in voicing, hence constituted phonotactically legal clusters in French. Care was taken to maintain the same volume, pitch, and intonation contour on all syllable pairs (rising on the test syllable and falling on the context). Syllables were then cut out using Cool Edit 2000.

From these raw materials, test syllables were constructed as follows: For the allophonic contrast, one token of [aʁ] and one of [aχ] was selected. The vowel portion of [aʁ] was replaced by that of [aχ], in order to ensure that the contrast only concerned the fricative portion. The fricative portions were then trimmed to 130ms in both syllables. The same procedure was applied for syllables with the vowels [i] and [u]. As to the phonemic contrasts, two tokens per syllable were simply selected from among the recordings. One token of each context syllable was likewise selected. Finally, disyllabic 'phrases' were constructed by appending all context syllables to all syllables constructed for the phonemic and the allophonic contrast.

[1] The assimilation rule applies both within and across words; although its maximum domain of application is not exactly known, the phonological phrase is a likely candidate.

2.2.2 Procedure

The experiment started with training during which subjects heard five pairs of two-word phrases in French made up of monosyllabic words, e.g. *veuf turc – veuf riche* ('Turkish widower' – 'rich widower'). They were asked to determine whether the first words of the two phrases were identical or not. Answers were given via 'same' and 'different' response buttons. During the training, feedback was provided.

Perception of the allophonic and phonemic contrasts was then assesed in two blocks, one in which subjects heard the contrasts in syllables in isolation, and one where these were presented with a following context syllable.

For the block with isolated syllables, subjects were told that they would hear two monosyllabic words in a foreign language and that they should say if they were identical or different. (The stimuli were described as words to try to prevent subjects from performing the task entirely acoustically.) Trials consisted of two syllables presented with an ISI of 500ms. They were defined by crossing the following parameters: contrast (allophonic *vs*. phonemic), vowel ([a] *vs*. [i] vs. [u]), first consonant ([ʁ] *vs*. [χ] or [m] *vs*. [n]), and second consonant ([ʁ] *vs*. [χ] or [m] *vs*. [n]). The 24 (2 x 3 x 2 x 2) trials thus defined were repeated 4 times, for a total of 96 trials. Overall, half of the trials consisted of identical syllables and half of different ones. For the trials with the phonemic contrast, tokens were chosen randomly.

Perception in phonological context was tested by presenting subjects with pairs of [VC.CV] sequences. They were told they would hear two short sentences constituted of two monosyllabic words in a foreign language and that as in the training, they should indicate whether the first words of both sentences were identical or different. There were again 96 trials, following the design outlined above, but with the difference that each test syllable was followed by a context syllable. Half of the trials for the allophonic contrast contained sequences with a consonant cluster that agreed in voicing and hence were phonotactically legal (e.g. [aʁdo]-[aχsa]), whilst the other half of the trials contained sequences with a cluster disagreeing in voicing and were phonotactically illegal (e.g. [iχbe]-[iχzu]). For the phonemic contrast, all sequences were phonotactically legal, since nasal consonants are freely followed by both voiced and unvoiced obstruents in French. Context syllables were chosen randomly, respecting the design outlined above and with the additional constraints that within trials, they had to be different with respect to both their vowel and their consonant, and that their vowels had to differ from the one in the corresponding test syllable.

In both experimental blocks, trials were presented randomly, and subjects replied by pressing one of two response buttons without getting feedback. A pause of 2000msec separated trials from one another.

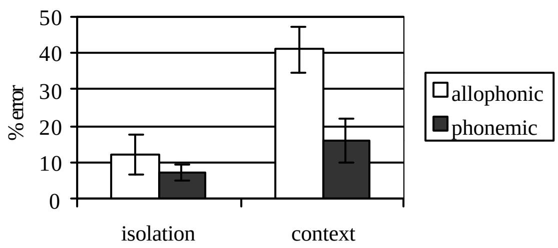
2.1.3 Participants

We tested twelve monolingual French native speakers. We excluded individuals who fluently spoke languages containing [χ] or the phonetically similar unvoiced velar fricative [x] (e.g. Spanish, German, Dutch, and varieties of modern Arabic). The participants were evenly distributed into two groups, one heard the isolated syllables first, the other heard the syllables in context first.

2.2 Results

Results are summarized in Figure I.

Figure I



Since there was no main effect of group ($p>.1$) nor any interaction with the experimental factors ($p>.1$), results from both groups were combined for all following analyses.

An ANOVA with factors context and contrast was then carried out, finding an effect of context ($F(1,11)=55.8$; $p<.02$) – discrimination is better without than with phonological context – and contrast ($F(1,11)=32.4$; $p<.02$) – discrimination of the phonemic contrast is better than discrimination of the allophonic contrast. The interaction between context and contrast was significant ($F(1,11)= 28.4$; $p<.02$), as the difference between phonemic and allophonic items was significant in the condition with phonological context ($F(1,11)=62.2$; $p<.02$) but not in the condition without context ($p>.1$). No other effect or interaction was significant.

For the condition with context, we also calculated the percentages of errors in phonotactically legal (40%) and illegal (43%) sentences. A restricted analysis yielded no significant effect for legal versus illegal phonotactics ($F(1,11)= 0.65$; $p>.1$). No other effect or interaction was significant.

2.3 Discussion

The results of this experiment show that [ʁ] and [χ] are acoustically different enough to be discriminated by French native speakers at a high level of performance when they are presented in isolation. Indeed, performance on the allophonic contrast is not significantly different from performance on the phonemic contrast. When embedded in a phonological context, however, the two types of contrast differ dramatically, in that errors on the allophonic contrast now rise to a very high level (although still significantly different from chance, $p<.02$), while the error rate on the phonemic contrast stays low.

What is the mechanism that underlies the increased difficulty of the allophonic contrast in context? A low-level masking explanation is unlikely. If the addition of the context syllable resulted in low-level degradation in the perceptibility of the coda consonant of the test syllable, we should have observed a similar impairment of the performance for the phonemic contrast. Alternative explanations all imply that the allophonic contrasts are processed in a 'special' way. We discuss two such explanations.

The first interpretation is that allophonic contrasts are represented as two distinct phonological categories, but that the processing system performs some kind of 'inverse phonology' mapping, thereby undoing the effect of phonological rules and recoding the segments accordingly. Under this interpretation, both the voiced [ʁ] followed by a voiced consonant and the unvoiced [χ] followed by an unvoiced consonant will be recoded as a single underlying phoneme /r/, making discrimination impossible. This predicts that when the segments are presented in a context that does not correspond to the assimilation rule (with a consonant disagreeing in voicing), the performance should go back to that of the presentation in isolation. We found that this was not the case, allowing us to discard this interpretation.

The second interpretation is that allophonic variants are not represented as two distinct categories but rather as a single phonological category. Under this interpretation, the discrimination of allophones is similar to the discrimination of non-native contrasts that are mapped onto the same native category, like the English /r/ and /l/ for Japanese speakers (Goto 1971). There is one caveat, though, because our participants were very good at perceiving the allophonic contrast when presented in isolation, whereas non-native contrasts are usually hard to perceive even out of context. Yet, previous research has shown that when within-category contrasts or non-native contrasts are presented out of context and/or within a very short time interval, they are easier to discriminate than within context, and/or after some time (Pisoni 1973, Werker & Tees 1984b). This facilitative effect of the absence of context can be interpreted by increased access to phonetic or acoustic correlates of the contrast, which are transitorily accessible in some kind of low-level acoustic buffer. Further research is needed to determine whether we are dealing with the same kind of phenomenon here.

Whatever the explanation for the present experiment, it clearly shows that allophonic contrasts are difficult to discriminate by French listeners when they appear embedded in a phonological context. In the next experiment, we investigate whether such difficulties can be alleviated or worsened by exposing participants with an artificial 'language' which distributional properties either favour or disfavour the creation of two distinct phonological categories for these sounds.

3. Experiment 2: Acquisition of Phoneme Categories

Maye & Gerken (2000) showed that the perception of allophonic contrasts can be modified after exposure to an artificial language containing tokens of the allophones with a certain statistical distribution. They tested the perception of the allophonic contrast between voiced [d] and the voiceless unaspirated [t] in English. Adult native speakers of American English were exposed to either a monomodal or a bimodal distribution of tokens lying on a continuum between these two sounds. After exposure, subjects in the bimodal group performed significantly better than those in the monomodal group in a discrimination task, suggesting that the former but not the latter had constructed two separate categories.

In our second experiment, we test, firstly, whether the same effect arises when the allophonic contrast is presented within context. To this aim, we use a continuum between [ʁ] and [χ] and expose French subjects to either a monomodal or a bimodal distribution of syllables that end in such tokens and are followed by a context syllable. We predict that subjects in the first group create a single category, whereas those in the second group create two categories; consequently, discrimination of the contrast should be better for subjects in the second group than for those in the first group.

Secondly, we examine whether the bias to create two categories in the presence of a bimodal distribution is overridden if the two segments have complementary distributions. To this aim, we expose a third group of subjects to a bimodal distribution of the continuum with the additional condition that the initial obstruent in the context syllable agrees in voicing with the preceding uvular fricative. That is, tokens at the [ʁ]-side of the continuum are followed by voiced obstruents only and those at the [χ]-side by unvoiced ones. According to the hypothesis of Peperkamp & Dupoux (2002) concerning the acquisition of phoneme categories, subjects are sensitive to these complementary distributions. We thus predict that subjects in this group create a single, abstract, phoneme category, and hence discriminate less well between the two segments than those in the second group. The differences between the three groups and our predictions are summarized in Table I.

Table I

	Token statistics	Complementary distribution	Predictions
1	monomodal	no	1 category
2	bimodal	no	2 categories
3	bimodal	yes	1 category

3.1 Method
3.2.1 Stimuli

We used the test syllables for the allophonic contrast and the context syllables created for Experiment 1. For each of the three pairs [Vʁ] – [Vχ], we created an 8-step continuum by cross-splicing the tail of the voiceless fricative onto the beginning of the voiced fricative. The relative duration of the voiced fricative portion compared to that of the unvoiced fricative portion along the continua was 100%, 85.8%, 71.5%, 57.2%, 42.8%, 28.5%, 14.2%, and 0%.

3.2.2 Procedure

The experiment consisted of four parts, i.e. training, pretest, exposure, and posttest. There were three subject groups, called Monomodal, Bimodal, and Bimodal + Assimilation; only the exposure was different for these groups.

The experiment started with a training that was the same as in Experiment 1. The pretest then assessed capacity to discriminate the [ʁ-χ] contrast in context. The task and instructions were the same as for the context block of Experiment 1. There were 48 trials, consisting of pairs of [VC.CV] sequences presented with an ISI of 500ms. Half of the trials were as described in Experiment 1 and contained either two agreeing or two disagreeing clusters. In the other half, one sequence contained a cluster that agrees in voicing and one that disagrees. These trials were defined by crossing the following parameters: vowel in the test syllables ([a] *vs.* [i] *vs.* [u]), order of the clusters (agreeing-disagreeing *vs.* disagreeing-agreeing), type of agreeing cluster (voiced or unvoiced), and type of disagreeing cluster (voiced-unvoiced *vs.* unvoiced-voiced). The choice of the obstruent and the vowel in the context syllables of these trials were subject to the same additional conditions as in Experiment 1. Overall, half of the trials had identical initial syllables, and half had different ones. During this part, only the endpoint stimuli of the test syllables were used.

Next, subjects listened to a list of [VC.CV] sequences presented with an ISI of 1000ms. To maintain attention, they were asked to check a box on an answer sheet after each sequence. There were three lists that differed in the type of distributional curve concerning the [ʁ-χ] continuum, as well as in the phonological distributions of [ʁ] and [χ]. As to the distribution of tokens along the continuum, Figure II (adapted from Maye & Gerken (2000)) shows the number of occurrences of the 8 items in the monomodal and bimodal curves.

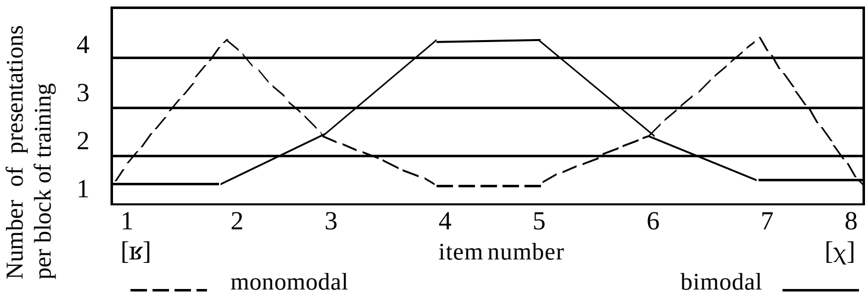


Figure II

Both curves correspond to 16 tokens. Subjects in the group Monomodal were exposed to the monomodal curve and those in the groups Bimodal and Bimodal + Assimilation to the bimodal one. As to the phonological distribution, for subjects in the groups Monomodal and Bimodal, context syllables were chosen randomly, whilst for those in the group Bimodal + Assimilation, context syllables were chosen randomly, with the constraint that tokens at the [ʁ]-side of the continuum were followed by syllables starting with a voiced obstruents only and those at the [χ]-side by syllables starting with unvoiced ones.

There were 24 blocks of 16 sequences each, for a total of 384 sequences and nine minutes of listening time. Within each block, context syllables were chosen as described above, and sequences were presented randomly. Subjects could take a short break halfway the exposure.

After exposure, subjects' discrimination of the [ʁ-χ] contrast was tested again. The task and instructions of this posttest were identical to those of the pretest. It contained the same 48 trials with the endpoints of the continuum, plus an additional 24 trials in which the uvular fricative in both sequences was a new token, 4½, lying at the middle of the continuum (hence containing the first 50% of [ʁ] and the last 50% of [χ]). These trials were defined according to the same symmetrical design as the other 48. Overall, subjects heard equal numbers of tokens 1, 8 and 4½. This distribution being incompatible with that of the exposure for both the monomodal group and the two bimodal groups, a possible effect of retraining should equally affect the three groups. Responses for the 24 trials with token 4½ were not taken into account for scoring. All trials were presented randomly.

3.1.3 Participants

60 Subjects with the same linguistic profile as in Experiment 1 were tested, i.e. 20 per group.

3.2 Results

Results are summarized in Figure III.

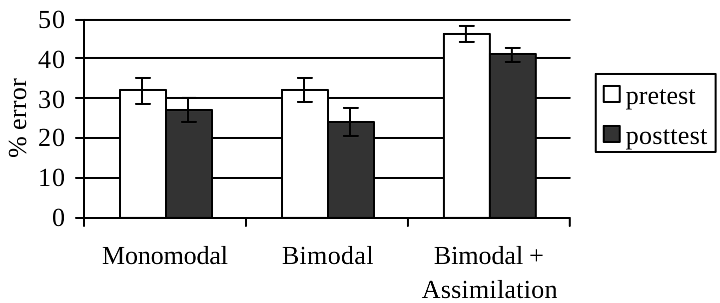


Figure III

A global ANOVA on the error percentage was run, with block (pre- *vs*. posttest) and group (Monomodal *vs*. Bimodal *vs*. Bimodal + Assimilation) as experimental factors. There was a significant effect of both block ($F(1,57)=13.5$; $p<.001$) and group ($F(2,57)=7.46$; $p<.001$), but the interaction between these two factors did not reach significance ($p>.1$). As can be seen in Figure III, the difference in performance across the groups was present even in the pretest. This difference was due to the group Bimodal + Assimilation which yielded significantly more errors in the pretest than the other two groups. The group Bimodal + Assimilation was tested after the first two groups were run, and it is possible that the participants in this last sample were simply less attentive to the task at hand. Regarding the effect of block, we found that it's amplitude was not the same across groups. Planned comparisons revealed that the difference between pre- and posttest was significant in the group Bimodal ($F(1,19)=6.92$; $p<.02$) but only marginal in the groups Monomodal ($F(1,19)=3.05$; $.1 > p > .05$) and Bimodal + Assimilation ($F(1,19)=4.04$; $.1 > p > .05$).

3.3 Discussion

Our results show a small but overall significant decrease in error rates in the posttest compared to the pretest. This decrease was numerically largest and significant in the group Bimodal, and smaller and only marginal in the other two groups. The difference in the amplitude of the decrease, however, was not large enough to yield a significant interaction across groups.

These results are in accordance with our hypothesis that statistical learning does not only rely upon a local prototype extraction mechanism, but takes into account the context in which segments are realized as well; in particular, it is sensitive to complementary distributions (Peperkamp & Dupoux 2002). In fact, the group Bimodal yielded the strongest amount of learning, and the groups Monomodal and Bimodal + Assimilation yielded equivalent small and only marginally significant amounts of learning. However, in the absence of a significant interaction between block and group it is difficult to assess whether these differences are real or result from chance. The lack of interaction is due to two features of our results. First, in the group with the strongest learning, i.e. Bimodal, the change in performance before and after exposure is numerically very small, only 8.4%. This is much smaller than the difference expected from Maye & Gerken's (2000) results, which is of the order of 20%. Second, there was a small and marginally significant amount of learning in the group where there should not have been any, i.e. Monomodal. As a result, the groups Bimodal and Monomodal do not differ statistically, which constitutes a failure to replicate Maye & Gerken (2000). Note that this failure was obtained even though our exposure contained twice as many items.

There are several explanations for the weakness of our results. The first one is similar to the explanation we provided for the context effect of Experiment 1. That is, the presentation of allophones outside context as in Maye & Gerken's experiment allows participants to operate at the acoustic/phonetic level and construct new prototypes on the basis of the statistical distribution of the tokens. By contrast, the introduction of context may make the acoustic/phonetic level much less available, thus allowing only to operate at the phonological level, where the distinction between allophones is already lost. Hence, context would - at least in adults having a single phonological category for the allophones - significantly slow down statistical learning. A second explanation is that our stimuli are more complex and varied than the ones used by Maye & Gerken. That is, the multiple context syllables might have distracted participants or diluted their attention and hence weakened or slowed down learning. A final explanation rests with the within-participant design that we used, which forced us to introduce a pretest. In contrast, Maye & Gerken didn't use a pretest, and directly compared the posttests of groups subjected to different types of exposure. The presence of the pretest could already function as a short but effective training procedure and focus the participants' attention on the particular contrast under study. Such an attentional focus would make the subsequent exposure less effective, because the participants would be less naive and already know that there are two different 'r' sounds to look for.

These different explanations are not mutually exclusive. In order to further buttress our hypothesis, we therefore need to conduct more research aiming at increasing the amount of learning obtained with the particular set of stimuli that we use. Several directions are possible: testing a population of adults in a language that does not use the two segments as allophones of one another, reducing the amount of phonological variability in the word endings, removing the pretest, or increasing the amount of exposure even more.

4. Conclusion

We have tested the perception of an allophonic [ʁ-χ] contrast in French listeners in two experiments. In Experiment 1, we found that the allophonic contrast is very well discriminated when presented in isolation, but very hard to discriminate when embedded in context. In Experiment 2, using a training procedure with stimuli in context, we found only a numerically small amount of learning overall and limited evidence that the statistical and phonological distribution of the tokens influence the amount of learning. These results are encouraging, but a substantial increase in the power of the training procedure is needed before definitive conclusions can be drawn regarding the mechanisms underlying the acquisition of phoneme categories. These results also suggests that phonetic processing and learning could behave quite differently when studied with tokens surrounded by silence as opposed to tokens embedded within a larger phonological context.

Ultimately, we should test infants with the training procedure of Experiment 2 to examine whether they can acquire abstract phoneme categories on the basis of distributional information.

References

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9: 317-23.

Hallé, P., J. Segui, U. Frauenfelder & C. Meunier (1998) Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24: 592-608.

Kahn, D. (1976) *Syllable-based Generalizations in English Phonology*. Bloomington: IULC.

Kim, N.-K. (1990) Korean. In: B. Comrie (ed.) *The World's Major Languages*. Oxford: Oxford University Press: 881-898.

Kuhl, P., J. Andruski, I. Chistovich, L. Chistovich, E. Kozhevnikova, V. Ryskina, E. Stolyarova, U. Sundberg & F. Lacerda (1997) Cross-language analysis of phonetic units in language addressed to infants [see comments]. *Science* 277: 684-6.

Kuhl, P., K. Williams, F. Lacerda, K. Stevens & B. Lindblom (1992) Linguistic experience alters phonetic perception in infants by 6 months of age [see comments]. *Science* 255: 606-8.

Lahiri, A. & W. Marslen-Wilson (1991) The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38: 245-94.

Maye, J. & L. Gerken (2000) *Learning phonemes without minimal pairs*. Proceedings of the 24th Annual Boston University Conference on Language Development. Boston.522-533.

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

Pegg, J. & J. Werker (1997) Adult and infant perception of two English phones. *Journal of the Acoustical Society of America* 102: 3742-53.

Peperkamp, S. & E. Dupoux (2002) Coping with phonological variation in early lexical acquisition. In: I. Lasser (ed.) *The Process of Language Acquisition*. Frankfurt: Peter Lang: 359-385.

Pisoni, D. (1973) Auditory and phonetic memory in codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 13: 253-260.

Polka, L. & J. Werker (1994) Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421-35.

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

Werker, J. & R. Tees (1984b) Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75: 1866-78.

Whalen, D., C. Best & J. Irwin (1997) Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25: 501-528.