UCL Working Papers in Linguistics 10 (1998)

Yorùbá babies and unchained melody[*]

PHIL HARRISON

Abstract

Earlier work on the perception of stop contrasts by infants and of lexical tone by Yorùbá adults sets the context for the results presented here, which demonstrate the sensitivity of six- to eight-month old Yorùbá babies to the tonal contrasts of their native language. Like adult speakers, they discriminate high tone from other tones in a single nuclear domain, but fail to discriminate mid from low tone. English-acquiring controls remain completely uninterested in such distinctions. These findings are readily aligned with the characterisation of melodic objects as privative elements, and so lend psycholinguistic credence to this model and provide evidence of the acquisition of laryngeal phonological contrasts in the first year of life.

1 Introduction

> 'Acquisition issues cannot be introduced in a theoretical vacuum. An acquisition model must be paired with a model of the component that is being acquired... I would regard any linguistic theory with some suspicion if its paired acquisition model were unduly complex, not to say impossible' (Kaye 1997).

Melodic primes can be represented as individual abstracts (elements). These are independent of prosodic structure, with which they interact via relationships which may be expressed in terms of licensing. (Kaye, Lowenstamm & Vergnaud 1985; Harris 1994;

[*]Thanks to UCL Graduate School, and to the department of Phonetics and Linguistics for providing funding for the expenses associated with the experiments reported here. An intellectual debt is owed to John Harris for his usual unusually helpful comments. Thanks also to Steve Nevard for building the reinforcer circuitry and nailing down the penguins' feet, to the staff at Wolfson House for allowing some of their valuable space to be used up, and in particular to those staff and students kind enough to assist with the experiments. Finally, to the Yorùba and English infants and to their parents who were so generous with their time and effort.

Harris & Lindsey 1995) Acquisition of prosody and melody must therefore be independent: ‘acquisition of phonology’ is formally meaningless. With this in mind, the sensitivity of infants’ perception to the contrasts of the language-specific melodic system they will later employ may evince the acquisition of ‘raw’ melody (Harrison 1995). An investigation into the perception of tonal melody by infants is here described with the express aim of setting this perception in a phonological context.

2 Categorical perception of stop contrasts by infants

The results of some experiments in the categorical perception[1] of stop consonants by infants show a sensitivity to exactly those boundaries that will become relevant to the adult phonology. Eimas, Siqueland, Jusczyk & Vigorito (1971) detected a Voice Onset Time[2] (VOT) discriminatory boundary at around +25ms for bilabial stops in English-acquiring infants as young as one month old. This coincides with the VOT boundary which cues the fortis/lenis distinction in the steady-state language.

Not all languages that utilise ‘voicing’ distinctions cued by VOT values[3] put the categorical boundary in the same place. Spanish, for instance, has a ‘voiced/voiceless’ demarcation at about 0ms VOT. In view of this, it becomes possible to investigate how linguistic input may alter such perception during infancy. Eilers, Gavin & Wilson (1979), using the Visually Reinforced Infant Speech Discrimination (VRISD) paradigm[4], reported a cross-linguistic study of 6-8 month old infants acquiring Spanish and English. Their results showed that while the English children were more sensitive to the English system, the Spanish children were statistically more able to perceive *both* the foreign- and the native-language boundaries. Thus the Spanish children at this age are in the

[1] The imposition of perceptual category boundaries at scientifically arbitrary values within a continuum.

[2] The time difference between the release of closure and the onset of vocal-fold activity in stops.

[3] In the mature language, VOT is not the only acoustic cue for ‘voicing’, the audibility of formant transitions being at least as perceptually important for some speakers and for some distinctions. VOT remains, however, robustly relevant to these perceptual discriminations.

[4] Described in more detail in §3.1 (below).

process of transferring systemic import from the 'factory setting'[5] to the native categorical value.

The traditional fortis/lenis bifurcation of English stop consonants is characterised in element-theoretic terms as a distinction between segments containing the laryngeal element [H] (the aspirated series) and one which lacks this element (the non-aspirated series).[6] The lenis series is therefore neutral as far as laryngeal specification is concerned, and in this respect identical to the southern French, Greek and Spanish 'voiceless' stops. In these languages, the phonological distinction is between the neutral series and the 'fully voiced' set, which contains the element [L], whose presence can be betrayed by the assimilation of voicing to adjacent positions, a phenomenon which never occurs from English lenis stops (Harris 1994: 133-137).

Now laryngeal elements [L] and [H] are also proposed to be potentially componential in nuclear positions where they are the active tonal elements in lexical tone languages (Kaye, Lowenstamm & Vergnaud 1990).[7] Given these acoustic *alter egos* of [H] and [L] in different prosodic positions, there is an opportunity to see if the demonstrated perception of VOT by six month old children is paralleled by a similar sensitivity to tonal contrasts, providing us with one test of a maximally simple acquisitional hypothesis, namely that perception at 6-8 months of age is already being transduced into phonological information having systemic import.

3 Infant perception experiments

3.0 Yorùbá possesses a three-way lexical tonal contrast: H, M and L. It has been demonstrated (Harrison 1996) that adult Yorùbá speakers may perceive the 'H/other' contrast for tone without needing a context that stretches beyond a single nuclear domain. If we can demonstrate that this attuned perception of tonal melody is also in place by six to eight months of age, then we achieve some chronological justification for

---

[5] Also possessed by chinchilla: see Kuhl & Miller (1975).

[6] A brief definition of 'elements' as 'phonetically-interpretable melodic primes' will suffice for present purposes

[7] There are diachronic arguments for this identification: see Haudricourt (1954), Hyman (1975) and Blumstein (1991) for some discussion.

characterising infant perception as ‘phonological’. We will also have obtained circumstantal evidence for the identity of the elements underlying source contrasts in both stop and tonal manifestations. In addition, the adult subjects failed to discriminate L from M tone in isolate syllables: if the babies can be shown to perceive H from other tones, and also to share this lack of discrimination, we have a further parallel between adult and infant mentation.

3.1 Methodology

The Visually Reinforced Infant Speech Discrimination Paradigm (VRISD) can be used with subjects of six to eight months of age and, compared to other strategies for testing infant perception (high-amplitude-sucking or heart rate) allows significant results to be obtained from a comparatively small sample (Eilers, Wilson & Moore 1977, Kuhl 1993). The proposed apparatus and method are therefore adapted from the experiments reported in these papers, with an experimental site arranged as shown in Figure 1.

The paradigm comprises a *Training Stage* and a *Testing Stage*. The aim of the *Training Stage* is to enable the infant to connect any change in the acoustic stimulus from the speaker with the appearance of the visual reinforcer (henceforth VR). The VR is an animated toy animal or animals[8] in a plexiglass box, visible only when lit: this type of VR has previously been shown to be highly reinforcing for infants of six to eighteen months of age (Moore, Wilson & Thompson 1977). It requires a 45-degree head-turn by the infant for it to be visible. The infant is seated on its parent’s lap throughout.

[8] In these tests, pink and blue drumming penguins were used.

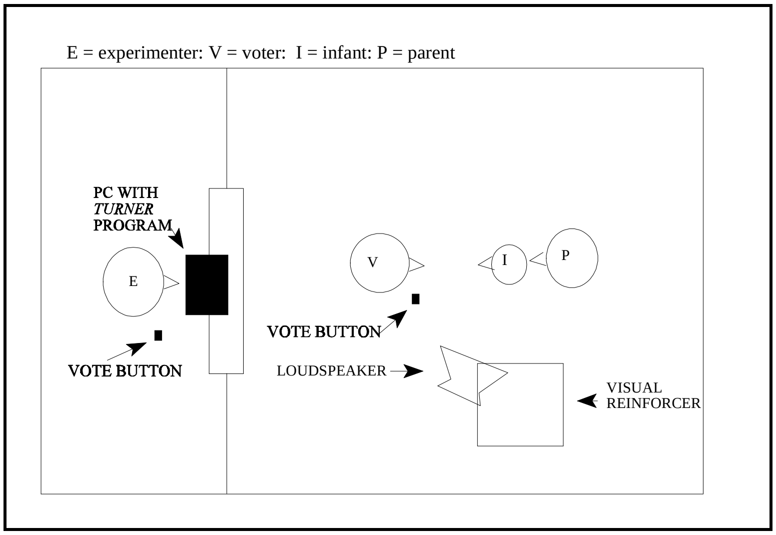

Figure 1: Experimental suite for infant speech perception testing.

During the *Training Stage*, a pair of two easily discriminable stimuli are chosen. These are loaded into the *Turner* program[9] in the PC, which replays the first at a regular interval of one or more seconds.[10] When the infant is judged by the experimenter (E) to be calm and attentive, the second stimulus is substituted for the first and the VR is turned on for a few seconds. This process is repeated until the infant turns to look after a change of stimulus but *before* the initiation of the VR. According to the papers cited above (Eilers, Wilson & Moore 1977, Kuhl 1993) most subjects turned at the first presentation of the new stimulus after two or three trials. Once this reaction has been established, the experiment can proceed to the *Testing Stage*.

We now choose stimuli whose discriminability we wish to assess. A pair of these are loaded into the *Turner* program. The first is played through the speaker at two-second intervals, as before. The infant is kept visually stimulated by the voter, who manipulates toys at the child's visual midline.The VR is now activated by the logic circuit illustrated

---

[9] Developed by Mark Huckvale at UCL.

[10] The interval is adjustable. Two seconds appeared to be the optimum value for keeping the infant attentive: this accords with earlier work (Kuhl 1979: 1670).

in Figure 2 (see Eilers, Wilson & Moore 1977: 769).

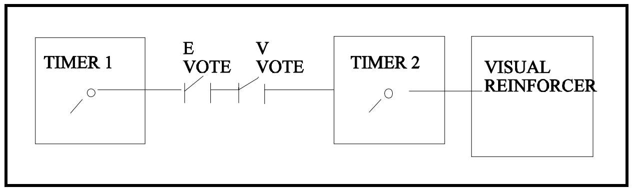

Figure 2: Logic circuit for activation of reinforcer during the *Testing Stage* of the experiment.

Timer 1 defines an interval - the *trial* - during which either a *Different* (D) or a *Control* (C) stimulus is presented to the infant. D is the second stimulus of the pair, and C is the first, hence a C option involves no change in the acoustic signal, while a D option involves a change. Each *trial* is of about four seconds duration. When the experimenter judges the child to be calm and attentive, (s)he initiates a *trial*. This command also activates a tone in the headphones of the voter who looks out for a head-turn by the infant.The *Turner* program decides which option (D or C) to choose, and randomises the choice. If a response (head-turn) is observed by either the experimenter or the voter, they depress their voting button. If both are depressed, the VR is activated for an interval defined by Timer 2. Ten *trials* take place for each pair of stimuli, the randomiser being adjusted so that: (i) there are five D and five C intervals, and (ii) no more than three D or three C intervals are consecutive.

During this procedure, both the voter and the parent wear headphones and listen to music to prevent their influencing the infant's response: the experimenter cannot hear the stimuli, or be seen by the infant.

The *Turner* program logs its choice of D or C *trial*, the timing of a positive decision (button-pressing) by either the experimenter or the voter, and the activation of the VR. After ten trials, it alerts the experimenter that the requisite amount of data has been collected for the stimulus-pair under investigation.

The proportion of responses during D intervals can then be compared with the proportion of responses during C intervals. For *six* trials (3 D and 3 C), statistical significance was said to be achieved in the Eilers Wilson & Moore (1977) experiment if the following results were observed:

(i) A head turn during all three D intervals and no C intervals: ($p < 0.01$)
(ii) A head turn during all three D intervals and one C interval: ($p < 0.05$)
(iii) A head turn during two D intervals and no C intervals: ($p < 0.05$)

These criteria are based on a *z*-test for proportion taken from Bruning & Kintz (1968). Aslin and Pisoni (1980) have questioned the accuracy of this analysis, and propose that at least ten trials are needed to achieve significance. Eilers Gavin & Wilson (1980) have replied in detail to Aslin & Pisoni's critique, but in the present experiment we sidestep this issue by utilising five D and five C intervals for any pair of stimuli used in the *Testing Stage*, and we claim significance only if 90% accuracy (nine out of ten *trials* with an appropriate response) is achieved.

3.2 Subject suitability

The subjects, for both the preliminary and the main tests, are divided into two groups. The first group are infants of six to eight months who are exposed to the Yorùbá language at home, and the second group are infants who are not. This second group come from either a monoglot English-speaking environment or from one in which, though other languages are sometimes spoken, no lexical tone language is used. Since all the babies live in London, they are all to a greater or lesser extent exposed to English. This is fortunately irrelevant to the experiment, because we are simply probing the effect of exposure to lexical tones in the first months of life (pre- and post-natal), and we propose to use isolated syllables as stimuli: the fact that both groups should be tuned in to some degree to the intonation of English should not bias our results. The sample size is small, but is similar to that which has previously been found to be adequate for this test, i.e. six normally developing infants of the requisite age in each group, who at the time of testing are not suffering from colds or middle-ear dysfunction (Eilers, Wilson & Moore 1977: 768). Introductory letters were then sent to parents of suitable infants, who were contacted by phone after a week or two to see if they would be willing to allow their children to be tested. Tests were carried out in a recording studio within the department of Phonetics & Liguistics at University College London. The testing took around eighteen months to complete, as parents within the Nigerian community in London, the only potential source of Yorùbá infants, proved reluctant to bring their children to be tested for reasons entirely beyond the scope of this report.

3.3 Choice of stimuli

Stimuli for the *Training Stage* of the experiment have the primary function of enabling the baby to associate a change in the acoustic signal with the appearance of the VR. In previous tests, amplitude has been used as a training parameter: Eilers, Wilson & Moore (1977: 773) presented the first stimulus at 50 dBSPL, and the second at 65 dBSPL. After obtaining a head-turn, the dB difference between the stimuli was reduced by 5dB steps. For the present tests, the first training pair were simply two highly discriminable Yorùbá words, low-toned [bà] (*perch*) and high-toned [kí] (*greet*). These were used both for controls and for the target Yorùbá infants. These two syllables differ along so many acoustic parameters that they should be easily discriminated by any trainable infant.

In certain cases attempts were also made to train infants on stimuli which contrast solely by virtue of a wide pitch differential (175/225 Hz). The success or otherwise of this procedure has implications for the mode of perception that the infant is demonstrating: see results in §3.4 below. These *Training Stage* pair of stimuli are additionally useful because they can potentially be used to retest subjects if a lack of ability to discriminate the *Testing Stage* stimuli emerges from results. Their status as *words*, i.e. syntactically specified objects with lexical significance, is irrelevant as there is no evidence that a child's linguistic development at six to eight months has reached a stage where such connections are integrated.

Similarly, although the stimuli for the *Testing Stage* must conform to the phonological constraints of Yorùbá, their status as *words* (in the above sense) is quite insignificant. For our purposes, every pair should comprise two stimuli which differ only in fundamental frequency value, the crucial cue for lexical tone.

Synthetic tokens of the 'words' [kí], [ki] and [kì][11] (*greet, thick* and *praise*) were therefore paired and loaded into the *Turner* program for the infant experiments, pitch contour and all other potential perceptual cues save for Fx value having been factored out of these stimuli. Individual stimuli have an Fx value (at the onset of the vowel) of between 145 and 240 Hz, and vary by 5Hz. Higher-pitched signals appear from other acquisitional work to draw a generally more positive response from babies: a parent will often raise the voice-pitch at the start of an infant-directed utterance.[12] It was therefore

[11] The creation of these stimuli is reported in Harrison (1996). New stimuli at 230 Hz and 240 Hz were manufactured for use during the infant tests.

[12] Thanks to Adrian Fourcin (pc) for this information.

worth seeing if this turned out to be a form of 'noise' in this type of experiment. The stimuli were paired using frequency differentials of 10, 20 or 40 Hz. The pilot study, which aimed to achieve a focus for later trials (see §3.4 below), used all three differentials, and the main study 20 Hz only.[13] Reliable discrimination of any pair less than 10 Hz apart by any child was neither expected nor predicted off adult perception tests (Harrison 1996): such pairs might putatively be used as controls near Fx values which, once testing has begun, seem to be providing potentially interesting results for the more widely-spaced pairs.

3.4 Infant perception tests: results and discussion of pilot study

Four preliminary tests, two using a Yorùbá infant, and two using an (English) control, were carried out to see if any particular stimulus pairs should be prioritised for later tests. As the time that an infant remains testable is comparatively limited, it was desirable, from a practical point of view, to use any potentially significant stimulus pair comparatively early in the procedure. Adult Yorùbá perception tests had already alerted attention to a possible perceptual boundary at about 190/210 Hz[14] (Harrison 1996), so pairs at around these values would certainly be used. However, at this stage there was no way of knowing whether or not our subjects could even discriminate acoustic pitch in isolate syllables, a skill potentially quite independent of their linguistic acquisition.

The subjects were trained on the stimulus pair [bà] and [kí]: training appeared to succeed each time, though it usually took between nine and twelve training runs, rather than two as has been previously reported (Eilers, Wilson & Moore 1977: Kuhl 1993). Following training, a haphazard selection of pairs of stimuli with pitch differentials at 10, 20 and 40 Hz were presented. The results are given in Table 4.4 (below), where 'x' indicates a positive response.[15]

---

[13] With one or two minor exceptions: see §3.5 (below).

[14] This is a post-normalised value: the issue of normalisation in this context is discussed in Harrison (forthcoming).

[15] But not necessarily a statistically significant one: see following text. For the pilot study, an *indication* of interest by the child, for instance by getting the first few responses correct and then apparently becoming uninterested was sufficient to enable priorities to be set for the tests proper.

Hz differential	Yorùbá 1	Yorùbá 2	English 1	English 2
160/200	x	0	0	0
180/220	x	0	0	0
200/240	x	x	0	x
160/180	0	0	0	0
190/210	x	0	0	0
220/240	x	x	0	0
190/200	0	0	0	
200/210	0	0		
220/230	0	0		

Table 1: Results of pilot infant perception tests. 'x' indicates subject interested, '0' indicates a complete lack of interest, and a blank space means that the particular subject was not tested on this pair of stimuli.

In neither case did the English subject respond in an systematic way to any of the pitch-differentiated stimuli.Yorùbá infant 2 appears similarly uninterested, apart from the in the cases of the highest 20Hz and 40Hz pairs, which may or may not have something to do with the relatively high pitch of these stimuli. Yorùbá infant 1 responded positively to pairs of 160/200, 180/220, 200/240, 190/210 Hz and 220/240 differential. This baby did not respond to changes of 10 Hz, or to a pair 160/180 Hz. The positive responses were not always statistically significant (i.e did not gain at least nine out of ten correct responses), but consideration of the results did concentrate attention for later testing on pairs of stimuli at a 20Hz distance. This decision was made on the basis of the following possible interpretations of these preliminary test results:

(i) The successful training runs indicated the ability to acoustically, rather than linguistically, discriminate A from B. The positive responses from one child to the pitch stimuli indicate that this was the only interested test participant. Linguistic acquisition is not being tested here. All infants can discriminate pitch in a likewise fashion: there is a tendency to cease to do so at around 10Hz of differential at this stage in life. As humans grow, they refine pitch discrimination down to the usual 4 or 5 Hz value for JND.

(ii) All the babies knew they were listening to language from the training stage. Yorùbá babies attend to pitch differences in isolate syllables, because they are at some stage in a process of attunement which will later map to lexical tone. English babies are not interested in this distinction.

(iii) All of the proposals in (ii) are true. In addition, Yorùbá babies interpret some pitch differences as phonological while others are simply pitch differences, and hence either less interesting or non-discriminable.

In other words, we are obliged to consider what *kind* of difference the infant is attending to. The notion of pitch-discrimination refinement in (i) above may be true, but it does not really affect the investigation. Physical pitch discrimination may proceed as it will, but its use by the developing mind-brain for the interpretation of cognitive structures is a separate issue. This is underlined by the simple fact that pitch discrimination is adapted as effectively during musical enculturation as it is during linguistic acquisition. Rouget & Schwartz (1970) document a Sudanese musical tradition which divides the octave into seven equal intervals, none of which correspond to anything in the diatonic system familiar to Europeans. A musical grammar is acquired as surefootedly by the unconscious as is a linguistic one, and trained musicians will readily acknowledge the depths of their non-specialist enculturation: having spent a lifetime experimenting with musical form to a point which discomfited the establishment and occasionally caused riots, Stravinsky says ‘ I myself have no habit of anything oriental ... in the Orient I recognise myself as a barbarian’ (Stravinsky & Craft 1959: 26).

Humans parse acoustic signals as language, music or noise: these modalities relate independently to physical acoustic perception, and the first two are interpreted via a structural cognitive system.[16] So although our present tests utilise pitch perception, the degree of finesse displayed by the subjects in this ability is only crucial to the enquiry

[16] The existence of ambiguity in the perception of, for example, birdcalls (music or noise?) does not undermine the discrete nature of the three perceptual routines: it is just that the nature of some acoustic signals potentially renders them available to more than one interpretation.

if we demonstrate that it is seminal, in this particular context, to the acquisition of (a particular) language.

This brings us to the first assertion of (ii) above, i.e.'all the babies knew they were listening to language from the training stage'. This is hard to disagree with, unless one advanced the (untenable) suggestion that *no* language acquisition has taken place at six months of age[17], or the equally unsupported notion that at this age a child has a mature, language-specific phonology. The training stimuli are simply primal CV syllables. If these were not processed as 'language' by the developing mind/brain, nobody would acquire a language. Equally, though the syllables do not obey the phonological well-formedness constraints for English words (they possess only one nuclear position), it runs entirely counter to all acquisitional research to suggest that humans are monoglot after six months of life: so much 'language-like' acoustic information is encoded in the training stimuli that both English and Yorùbá infants, if they attend, must (unconsciously of course) consider that they are attending to language. Given this, we may be justified in proposing a linguistic component to a subject's subsequent attention to the *Testing Stage* stimuli, and to regard a positive response to those stimuli as the consequence of a generalisation of a trained ability.

The proposal was advanced in Harrison (1996) that there is a post-normalised categorical perceptual strategy for lexical tone. If this is correct, then we may claim support for the notion that a child is attending to the phonology of the lexical tone component of the target language if a tendency to perceive 'difference' at a certain pitch value were demonstrated. Yorùbá infant 1's results give us at the very least a hint that this investigation may be worth pursuing, and that our focus on the 200Hz pitch value could just be on target.

In addition, we take forward to a more systematic set of tests the need to be cautious about the following factors:

(i) An infant's attention span is comparatively short and relatively unpredictable, and this should be taken into account, as far as ethically possible, in an attempt to get conclusive results;

(ii) It is possible that infants attend to pitch differences without any linguistic significance;

[17] This idea is insupportable in the light of the assembled research programs cited in Harrison (forthcoming), one or two of which are referenced in §2 (above).

(iii) It may be possible for infants to generalise from a particular linguistically significant change to such changes in general, which could defocus their attention from the target parameter.

3.5 Infant perception tests: results and discussion of the main tests

Subjects for these tests were six infants who were living in a Yorùbá speaking environment, and six infants who were not.[18] The children were between 0:6:10 days and 0:7:25. Age difference within these limits has not been found to be a significant factor in the work we have hitherto considered: Eilers, Gavin & Wilson (1979: 15-16) use children aged between six and eight months in their stop-perception tests, and Moore, Wilson & Thompson (1977: 330-332) show that while younger children cannot be usefully tested using VRISD, children within these chronological limits show a uniformly positive response. The tests reported here set out to answer the questions in (1):

(1) (a) Could any of the children generalise from training on a linguistic pair that differed multidimensionally to one that differed in pitch only? (Our pilot study suggests this is true, at least for Yorùbá children.)

(b) Could any of the children be trained on a widely-different 'pitch-only' linguistic pair, and then use this training to narrow down to finer differentials? If so, is this pitch-perception only or is it linguistically significant?

(c) If a constant (20Hz) differential were maintained, would any of the children respond significantly differently to stimuli at different pitch specifications?

The subjects were therefore all trained using 'bà' and 'kí' stimuli (as in §3.4 above). In addition, an attempt was made to train three in each group using a pair of stimuli comprising two tokens of the 'ki' syllable which maintained a 50Hz differential, at 175 and 225 Hz. These six children, if the pitch training proved successful, proceeded to the *Testing Stage*: if not, an attempt was made to train them on 'bà' and 'kí' before

[18] The role of non-Yorùbá language input is discussed in §3.2 (above).

proceeding. *Training Stage* results are presented in Table 2. *Testing Stage* stimuli were paired tokens of the syllable ‘ki’ which maintained an Fx differential as set out below in Table 3, where the results for this latter stage are presented.

	Yorùbá children						English children					
	1	2	3	4	5	6	7	8	9	10	11	12
*(a)*175-225		x		x		0		0		0		0
(b) bà / kí	x		x		x	x	x	x	x	0	x	x

Table 2: Results of infant perception training using (a) pitch-only stimuli at 175 and 225 Hz, and (b) generalised stimuli (‘bà’ and ‘kí’). ‘x’ in a cell indicates successful training, ‘0' indicates no success, and a blank cell indicates that this particular training was not attempted (see text above).

Following his success on the 190/210 Hz pair, Yorùbá child 5 was tested on pairs 10Hz apart at 190/200, and 220/230 without achieving significant results. English child 8 was *tested* on the 175/225 pair, again without success. The blank cells in Table 3 are simply a result of having to abandon the experiment when the subject became fractious or inattentive. Some of the ‘0' cells do not represent a full ten *trials*: if no head-turn at all had occurred after seven or so, the particular stimulus pair was abandoned and a new pair used in the interests of getting as much usable data as possible from each subject. Some tests were also abandoned prematurely for practical reasons.

For these reasons, the results are hardly as conclusive as we could wish, but three interesting patterns do emerge from Table 3, even if they can chiefly be assessed impressionistically. The first is that only Yorùbá children trained successfully on pitch. The second is that, within our sample, it is only the Yorùbá children who achieve any level of statistical significance during the *Testing Stage* in their response to pitch differences in isolated syllables. The third pattern to notice is that within the Yorùbá group itself, it is only close to the pitch value predicted off the adult tests and the infant pilot study (indicated by the shaded cells) that statistical significance is achieved.[19] These regularities merit an explanation. We will consider them in the light of the questions in (1) above.

	Yorùbá children						English children					
Hz diff. for *ki*	1	2	3	4	5	6	7	8	9	10	11	12
140-160		0			0	0						0
160-180			0				0					
170-190		0		0	0		0				0	
180-200	0				0	0			0			0
190-210	0*	0*	x	x	x	0	0	0*	0	0	0	0
210-230		0	0	0	0	0			0*		0	0
220-240	0	0*										

Table 3: Results of infant perception tests using stimuli of 20Hz pitch differential: ‘x’ indicates successful perception at a statistically significant level, ‘0' failure, and ‘0 ‘ that, while statistical significance was not achieved, there were some successes (the children may have got the first few right and then apparently lost interest: they did not achieve the requisite 90%).*

Question 1(a) was answered positively for the Yorùbá children and negatively for the English controls. The target group did attend to (some) pitch-only stimuli in the *Testing*

[19] It is obvious from the distribution of blanks in the table that the number of tests performed were skewed toward this target 190/210 Hz pair. This was achieved by presenting this pair comparatively early in the *Testing Stage*, before the infant became fatigued or distracted. In practice, this pair was presented either first, second or third.

Stage, and if the proposal, advanced above, that *all* the babies knew they were listening to language is accepted, then we have some justification for advancing the idea that only the Yorùbá infants *knew that pitch was linguistic*. The *caveat* that non-linguistic pitch perception may be a source of 'noise' in this kind of test has already been aired: in this light, it is interesting that the results do hint at a tendency, possibly independent of the acquisition of a particular language, to attend to pitch differences at a higher register. This harks back to our earlier remarks about the raised pitch-range typically used by mothers to address their young infants (§3.3 above).

A precursor to the answer to the question posed in 1(b) is that the prediction that only the Yorùbá babies would be interested in pitch in the context of a single nuclear domain during the *Training Stage* has been borne out. The two infants who achieved success at a 50Hz differential did go on to discriminate at least one of the 20Hz pairs at a statistically significant level. The fact that the successfully discriminated 20Hz pair lay entirely within the range of the 50Hz pair could, in isolation, indicate a tuning of the perception of pitch outside of a linguistic context. However, the negative results gained from testing on the pair at 180/200Hz (and to some extent the same goes for 170/190Hz and 210/230Hz) mitigate against this conclusion. We have to acknowledge that it is possible that a mode of perception other than pitch-only is being utilised, and of course the focus of the present enquiry is to see if it is reasonable to characterise that perception as phonological.

Our strongest piece of evidence that this is so comes from the answer to question 1(c). For the 190/210Hz pair, and only for this pair, and only from within the target group, statistically significant results were derived. If we align this finding with the proposal that lexical-tone perception is 'post-normalised categorical' in nature, and that normalisation is an early (earlier)-acquired, and more peripherally processed extralinguistic capability, then it is possible to advance the notion that the simplest-available explanation for the results is that the Yorùbá infants are perceiving raw phonological melody. Above some boundary value that is acoustically around 200Hz in the present set of stimuli, and independent of acoustic perception, a different language-specific configuration obtains than that which exists below (about) 200Hz. The existence of only one such boundary suggests a binary opposition for the perception of tone, rather than the ternary set of phonological categories that is consensually argued to be present in the adult language. This is at least consistent with the assertion in Harrison (1996) that [L] is not present in Yorùbá nuclei, and that what has hitherto been considered as the perception of [L] is in fact the perception of a prosodic constituent (or, as a weaker assertion, the perception of an object that is not of the same phonological status as [H]).

Like the adult Yorùbá speakers, the infants, out of context, perceive the difference between [H] and Ø, but not between [L] and Ø.

A *chi-square* test can be applied to the total number of responses during trials for each linguistic group, comparing sucesses with failures. The total number of recorded responses was 297. A successful response is a head turn during a D-*trial* or no head turn during a C-*trial*. The null hypothesis is either a random response or no response at all, each one delivering an expected 50% success rate. Comparing responses to the 190/210 Hz pair with the responses to all other pairs, no significant difference is observable for the control group. However, the Yorùbá group's results produce a rejection of the null hypothesis at the $p < 0.005$ level of significance. A contrast is also obtained from comparing results for the 190/210 Hz pair with those for the next most often tested pair, that comprising stimuli at 210/230 Hz. The result for the target group in this case is still a rejection at $p < .005$, but for the control group $p > .05$ (<.1).[20]

An alternative analysis of these findings is presented in Figure 3, where the number of successful responses within each of the categories *190/210 Hz, 210/230 Hz* and *all other stimuli*[21] is expressed as a percentage of the total number of tests for that category.

[20] Chi-square results: all tables with 1 degree of freedom.

		χ^2 Values
Yorùbá group:	190/210 Hz pair *v* all other results	24.1
	190/210 Hz pair *v* 210/230 Hz pair	24.0
English group:	190/210 Hz pair *v* all other results	1.4
	190/210 Hz pair *v* 210/230 Hz pair	3.3

[21] i.e., all save the 190/210 Hz pair.

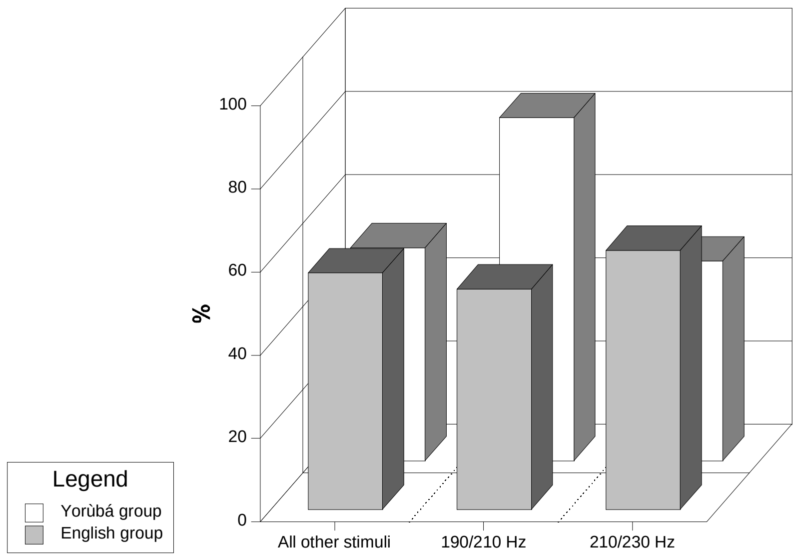

Figure 3: % successful responses for each group on (a) all stimuli except the 190/210 Hz pair, (b) the 190/210 Hz pair, and (c) the 210/230 Hz pair.

The story so far inspires both further empirical investigation and the search for support for these findings from independent sources. There are several extensions and adaptations to the present experimental paradigm that could be made to re-test our proposals, and to see in particular if the temptingly explicable result *vis-a-vis* phonological perception could be repeated. For example, there could be detectable acoustic properties present in the *natural* low tone utterances which cue the perception of [L]. This assertion is not, however, supported by the results reported in Bakare (1995). Bakare seeks to assess the relative usefulness of different acoustic cues in the perception of Yorùbá tones, and finds that Fx is the overwhelmingly significant cue to the perception of 'tonemes'.[22]

It may also be possible to glean results from similar tests carried out on older infants, or on those who are denizens of other linguistic environments. These are future possibilities.

[22] See Harrison (forthcoming) for a discussion of Bakare's work in the light of the present findings.

4 Summary and phonological context

The tests described here were substantially motivated by the objective of aligning lexical-tone perception in infancy with known findings on stop perception[23] insofar as both modes of perception could be seen to have developed in a chronologically parallel fashion. This would support the theory-internal prediction of the identity of the phonological objects mapped to by these perceptions. Evidence for this identity may have had to be derived in two stages: firstly, to show that infant pitch perception exists at the target age, and secondly that it is phonological. The contrasting results of the Yorùbá target group and the English control group has actually answered these questions simultaneously. Training subjects on linguistically viable stimuli, we have argued, must awaken the expectation of linguistic discrimations being possible from the testing stimuli: we owned that 'pure' pitch perception may act as 'noise', but the results appear to show that this is not so, at least in the present circumstances. The very fact that young Yorùbá (soon-to-be) speakers respond at all to isolate syllabes with a pitch difference while their English counterparts do not, together with the demonstrated bias of the Yorùbás to hear change at a single point in the frequency continuum, argues for language-specific influences. The theory-internal claim, therefore, that identical mental objects subtend tonal and source contrasts is supported so far *from a chronological point of view in acquisition*. But it has now become desirable, if we are to further shore up this claim, to show that Yorùbá tonology has one, rather than two, underlying lexical tones.

There are, in fact, a number of discrepancies in the patterning of tones in Yorùbá which do not accord with their being associated with two equal-status elements (for high and low tone) and a representation empty of source elements (for mid tone). Furthermore, these asymmetries extend to Yorùbá's dialectal cousins. In Harrison (forthcoming), it is argued that the most economic account possible of Yorùbá tonology is that only one element, the element [H], is manifested by tonal contrasts in nuclear positions. This element may interact with prosodic junctures to derive some aspects of the language that are diverse by dint of their phonetic measurement, and this story can be seen to correlate with the asymmetric melodic licensing capacity of nuclei in Yorùbá.

[23] Eimas, Siqueland, Jusczyk & Vigorito (1971); Eilers, Gavin & Wilson (1979).

References

Aslin R.N. & Pisoni D.B. (1980). Effects of early linguistic experience on speech discrimination by infants: a critique of Eilers, Gavin & Wilson (1979). *Child Development* 51, 107-112

Bakare C.A. (1995). Discrimination and Identification of Yoruba Tones: Perception Experiments and Acoustic Analysis. In *Language in Nigeria* (K. Owolabi, editor) Ibadan: Group Publishers

Blumstein S.E. (1991). The relation between phonetics and phonology. *Phonetica* 48, 108-119

Bruning J.L. & Kintz B.L. (1968). *Computational Handbook of Statistics* Glenview, IL: Scott Foresman

Eilers R.E., Gavin W., & Wilson W.R., (1979). Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development* 50, 14-18

Eilers R.E., Gavin W., & Wilson W.R., (1980). Effects of early linguistic experience on speech discrimination by infants: a reply. *Child Development* 51, 113-117

Eilers R.E., Wilson W.R., & Moore J.M. (1977). Development of changes in speech discrimination in infants. *Journal of Speech and Hearing Research* 20, 766-780

Eimas P., Siqueland E., Jusczyk P.W. & Vigorito J. (1971). Speech perception in infants. *Science* 171, 303-318

Harris J. (1994). *English Sound Structure* Oxford: Blackwell

Harris J. & Lindsey G. (1995). The elements of phonological representation. In *Frontiers in Phonology: Atoms, Structures, Derivations* (J.Durand & F.Katamba, editors) Harlow, Essex: Longman

Harrison P.A. (1995). Phonology and infants'perceptual abilities: asking the right questions. *UCL Working Papers in Linguistics* 7, 463-483

Harrison P.A. (1996). An experiment with tone. *UCL Working Papers in Linguistics* 8, 575-593

Harrison P.A. (forthcoming). *The acquisition of phonology in the first year of life* ms, UCL, London

Haudricourt A-G. (1954). De l'origine des tons en viêtnamien. *Journal Asiatique* 242, 68-82

Kaye J. (1997). Why this Article is not about the Acquisition of Phonology. *SOAS Working Papers in Linguistics and Phonetics* 7, 209-220

Kaye J., Lowenstamm J., & Vergnaud J-R., (1985). The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook* 2, 305-328

Kaye J., Lowenstamm J., & Vergnaud J-R., (1990). Constituent structure and government in phonology. *Phonology* 7, 193-232

Kuhl P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America* 66,1668-1679

Kuhl P.K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics* 21, 125-139

Kuhl P.K. & Miller J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190, 69-72

Moore J.M., Wilson W.R., & Thompson G. (1977). Visual reinforcement of head-turn responses in infants under 12 months of age. *Journal of Speech and Hearing Disorders* 42, 328-334

Rouget G. & Schwartz J. (1970). *Echanges et communications 1* Paris: Mouton

Stravinsky I. & Craft R. (1959). *Conversations with Igor Stravinsky* London: Faber & Faber