An experiment with tone^{*}

PHIL HARRISON

Abstract

It has been previously claimed that identical phonological objects are componential in phonation contrasts and lexical tone. It has also been shown that infants perceive language-specific phonation contrasts. We here investigate the perception of lexical tone by an adult speaker of Yoruba to allow us to both design stimuli and form hypotheses for infant testing. Two important theoretical points are discussed: firstly, that prosody and melody, which remain independent throughout adult life, must undergo separate developments; and secondly, that the insertion of identical melodic material into different structural architecture may lead to different phonetic outputs.

1 Introduction

Learnability is a perennially important issue for any grammatical model, and this applies to phonological acquisition no less than it does to any other aspect of a grammar. The aim of the following discussion is to bring into focus one potentially profitable area of enquiry into the acquisition of phonology in the first year of life. The phonetic literature abounds with tales of infant precocity. The perception of some phonation contrasts is present by one month of age (Eimas *et al.* 1971), and Kuhl (1992) indicates that language-specific prototypes for vowel perception have developed by the age of six months. On the prosodic front, Mehler *et al.* (1988) show that a child at the age of four days can distinguish its native language from timing and intonation cues alone. Though more and more 'phonetic' talents may continue to be ascribed to younger and younger children, neither a psychologist nor a linguist would want to limit the investigation to the depiction of these abilities. In a recent review article on progress in research in different areas of infant cognition, Karmiloff-Smith is right to remind us that '...we need to know not merely

^{*}Many thanks to John Harris for his helpful comments on an earlier version of this paper. Thanks also to Mark Huckvale, Steve Nevard and Andrew Simpson for their assistance with different aspects of the implementation of 'real' experiments. A very big thank-you is also due to Akin Oyetáde for his cooperation in the production of the Yoruba tokens and their subsequent use in perception tests.

whether such abilities are present, but how they are possible and what cognitive processes are involved' (Karmiloff-Smith 1995 p. 1307).

One obvious investigative strategy to employ in our search for the nature of these cognitive processes is to use the insights we already possess into adult mentation and seek evidence for these same processes in infancy. In this paper, we discuss some advances in the means of representing those phonological significances which have phonation and tone as their phonetic expressions (section 3), then following an investigation into the perceptual routine for lexical tone discrimination (section 4) we outline the predictions which are made by the results of this investigation about the mapping between infant phonetic perception and infant phonological development (section 5). We finish, in section 6, with some speculation about the universality of our account. However, since it becomes crucial to our argument that separate structures in the phonology undergo separate developments, and that the interaction of these structures may affect phonetic expression, it is worth our while first of all to support the notion that these structures remain phonologically discrete in the mature system.

2 Phonology as two interactive systems

I argued in Harrison (1995) that there is no logical reason to disallow the proposal that the acquisition of any part of a phonological system can be chronologically independent of the acquisition of any other modular component of the grammar, as long as we include certain assumptions in our theoretical model. The chief of these assumptions, detailed in Kaye, Lowenstamm & Vergnaud (1985 & 1990) and Harris (1994) among others, are:

- i) melodic primes can be represented as individually phonetically realisable abstracts (elements) and
- ii) that these are independent of prosodic structure, with which they interact via relationships expressed in terms of licensing.

Indeed, by adopting this theoretical position, we render the phrase 'acquisition of phonology' meaningless. If the mature system is represented as two interactive subsystems which remain independent, then we have no *option* but to propose that they are acquired separately. The alternative to this proposal is that they are acquired as one and differentiate later in life. Such an alternative would take some defending, and would require a general account of how it is possible to devolve two linguistic structural systems

from one supersystem. For example, the need to characterise an interface between morphosyntax and phonology is evident from the independence of morphosyntactic and phonological categories. One instance of this independence, taken from several cited in Spencer (1991 pp.42-43), is found in the assignment of stress in Czech. In this language, stress falls on the first syllable of a word. In the case of a monosyllabic preposition preceding a noun, though, stress is attracted by the preposition. If we use this part of the phonology to define the word boundaries, we deliver up the unlikely hypothesis that <na ten **stul**> ('onto that table') is three words, and <**na** stul> ('onto the/a table') is one. Explaining the *nature* of the interface between the two systems has of course been the subject of a good deal of scholarly endeavour, but its existence, in some form, is not in dispute.

So it is with our phonological model. While it has been convincingly proposed that within a single system simplex structures may bifurcate (for instance, in the acquisition of branching onsets), and while some form of combinability is germane to the account (for instance, in the characterisation of vocalic primes), the notion of prising apart two discrete systems in the course of acquisition seems to me entirely indefensible.

It is therefore our initial aim to review the uncontroversial notion that melody and prosody exhibit independence in adult language. We now briefly recapitulate some examples of the evidence for this assertion, and note that such evidence is drawn from three entirely independent data sources, and used to account for three entirely separate facts about language.

Firstly, Levelt (1992) takes the indications derived from spoonerisms, using these primarily to support the notion that the word-frames that are built in processing are phonological, and not lexical, in character. Thus (p.16) the utterance *peel like flaying* (for *feel like playing*) does not involve an exchange of consonants across the intervening word '*like*', but the swapping of the melodic content of the onsets of two adjacent phonological words [feelike] and [playing]. So far, this gives credence to the characterisation of syllabification as an independent process whereby melody is accommodated within an extant prosodic structure.

Levelt does point out that this neat modular account has been challenged, since there is robust evidence for a certain amount of lexical interference in phonological encoding: selection errors and encoding errors are not entirely independent (Dell & Reich 1981) and real words (as opposed to nonsense) are created by encoding errors more often than predicted by chance (Stemburger 1983, and others). But any misgivings about this type of evidence being indicative of a *purely* modular phonology in no way undermine its

usefulness in underlining the discrete nature of prosodic and melodic representations within the phonological component.

Our second example, drawn from Harris (1994 p.35), is of a diachronic change in English. A productive transition in phonetic realisation has taken place over time, and this is exemplified in (1):

STAGE 1	STAGE 2	STAGE 3	gloss
nıxt	ni:t	nait	'night'
rıxt	ri:t	rait	'right'
wext	we:t	weit	'weight'

The change between stage one and stage two is not simple fricative loss: the number of syllabic positions remains constant, and an instantiation of the widely attested process of compensatory lengthening means that the prosodic structure remains inviolate. Compensatory lengthening is formalised by Harris and others as a two-step procedure, involving first the delinking of a skeletal point, and second the relinking of that point to vocalic material. The relevance of these procedures to the present discussion is that neither of these steps involves the prosodic structure in any way, but confine their attentions to subsegmental objects.

The mirror image of this argument is to be found in our third piece of evidence for the independence of prosody and melody. Certain phonological constraints can be shown to ignore melodic material, and speak merely to the prosody.

The notion of prosodic minimality put forward in McCarthy & Prince (1986) states, among other things, that all English words must be bimoraic. Thus different phonetic surfaces may overlay structures which are invariant in this parameter. This notion, translated into the terms of the model we are here using, is illustrated in (2) for the English words 'city', 'sit', and 'say'.

(1)

0	R	0	R	0	R	0	R	0	R
ł	N		N		N		N		N
İ		Ì		Ì		İ		Ì	\setminus
х	х	х	х	х	х	х	х	х	хх
S	I	t	I	S	I	t		S	е і

These independently motivated analyses by Levelt, Harris, and McCarthy & Prince all have in common the demand that we recognise the discrete nature of prosody and melody in the mature system, and so reinforce our conviction that these systems must undergo separate developments.

So given even the feeblest appeal to modularity and using the structural primitives that we have outlined, there is no reason why developments in phonological acquisition should not take place long before the system is fully integrated. At the risk of double underlining something which is already patently obvious, it may be worth noting that even if we cleave to rival theoretical bases the dichotomy of melody and prosody persists in some guise. Within the system described in Coleman & Local (1992), the relation between phonetics and phonology is much more strictly demarcated than we have been assuming. Under no circumstances, in this model, is a phonological representation allowed to have an intrinsic phonetic interpretation. However, in their adoption of this system, Local and Lodge (1996) have to bifurcate the phonetic 'exponents'. 'The statement of phonetic exponents...has two formally distinct parts: temporal interpretation and parametric phonetic interpretation' (p. 94). In our terms, the first of these is clearly prosodic structure, and the second is melodic colour.

3 The representation of lexical tone and phonation

3.1 Introduction

When [H] and [L] show up in Onset positions, they are the active elements for phonation contrasts. Thus in elemental orthodoxy, [H] is present in aspirated stops and [L] in the fully voiced series (Kaye, Lowenstamm & Vergnaud 1990, Harris 1994). In Nuclei, tone bearing units are linked to either of these same elements, [H] for a high tone and [L] for a low. The neutral series of obstruents shares with mid tones the property of having no

(2)

(laryngeal) element. However, this analysis of lexical tone straight away begs some theoretical questions.

3.2 Asymmetries

There are languages such as Gujarati and Panjabi (for this latter possibility see the phonetic analysis advanced by Stuart-Smith (1996)), in which a four way contrast obtains for phonation. This allows us to fill in a predicted value in the paradigm, and propose that for 'breathy', or 'tonal', stops, *both* [H] and [L] are realised in Onset position (Harris 1994). But a theoretical explanation is required for the inability of *tone* bearing skeletal points to be linked to both [H] and [L]. The lack of phonological analyses which include this logical combination is wholly undesirable without a principled account of its absence. One possible partial explanation would be an appeal to tier geometry, wherein both elements could be fused onto a single tonal tier. This is analogous to the analysis of the English vowel system in which [I] and [U] are said to remain fused onto the 'colour' tier, accounting for the lack of rounded front vowels in the language. All the same, it is hardly felicitous for the theory that [H] and [L] should fail to dissociate from the same tier in *any* language.

Our second indication that something is wrong with the idea that the phonological reality underlying lexical tone discrimination can best be represented using two elements is to be found in a typological asymmetry. Some Bantu languages of southern Africa have been analysed using [H] only (Goldsmith 1990). Some tone languages of western Africa have, by contrast, been analysed using both elements [H] and [L] (Pulleyblank 1986). No language, however, has yet been analysed as using [L] only. Once again, this unpredicted asymmetry is so far theoretically inexplicable: no good reason has been advanced why [H] is to be 'favoured' over [L] in Nuclei.

3.3 Reduction exemplified

In recent years there have been attempts to reduce the number of elements, in order to achieve the laudable ambition of weakening the combinatorial power of the inventory. Some of this work has direct consequences for [L], and for what we have hitherto thought of as the phonetic reflexes of [L].

Cabrera-Abreu (1996), uses Pierrehumbert's theoretical framework which analyses intonational patterns as being reflexes of two monovalent primes ([H] and [L]) in association with prosodic boundaries. She then goes on to demonstrate that if this system is linked to the principles of directional licensing, licensing inheritance and the projection of nuclei, such principles being already independently present in the grammar, it is possible to reduce the inventory of primes in this context by one and reanalyse phrase level intonation patterns in English without [L]. [L] is targeted for extinction by Cabrera-Abreu for reasons ranging from its historical instability to its lack of attestation in the 'tone copying' processes participated in by [H].

Showing the same reductionist spirit as Cabrera-Abreu, Nasukawa (1995) argues that in an analysis of Japanese it is possible to conflate the element [L] with the element [N]. His account of the phenomena of Rendaku and 'Lyman's Law' (for a description of the facts and earlier accounts see Nasukawa 1995), hinges on the identification of the active element in Japanese voiced obstruents as *dependent* [L] and in Japanese nasals as *headed* [L]. Thus Rendaku is the attachment of [L] to the second term of a compound:

(3) [L] [ori] 'fold' + [kami] 'paper' = [ori gami] [L] [otoko] 'man' + [kokoro] 'heart' = [otoko gokoro]

Such attachment is prohibited by a following voiced obstruent in the same domain because this would create a symmetrical relationship of two dependent [L]s on the [L] tier and there can be no licensing relationship between two such objects.

(4) [otoko] 'man' + [kotoba] 'speech' = [otokokotoba] (*[otokogotoba])

The headed $[\underline{L}]$ in the nasal ([origami]) can contract an asymmetric relationship with the attached dependent [L], and so this problem does not arise.

Japanese voicing assimilation in compounds is also successfully analysed in Nasukawa's account as a manifestation of the same licensing constraint that he previously invoked: the potential sequence of two voiced obstruents that would result from [tob] 'fly' + [ta]

past (*[todda]) cannot occur. A head-dependent relationship must obtain on the [L] tier, and so [tonda] is the only possible output (Nasukawa 1995, pp. 5-6).

We should note that the identification of [L] with [N] in non-nuclear positions would also deliver a more symmetrical account of some active phonological processes in languages which share only a distant genealogy with Japanese.

Voicing assimilation in modern Greek is always the acquisition of positive phonation, regardless of direction, as can be seen from the following three examples.

(5)	$[tis]+[vrisis] \rightarrow [tizvrisis]$	'of the fountain'
	[enan]+[kseno] → [enaŋgzeno]	'a stranger'
	$[tus] + [neus] \rightarrow [tuzneus]$	'the news' (acc.)

Greek has fully-voiced and neutral obstruents, and no aspirated set: in this it is like French, in which such assimilation also takes place, and unlike most dialects of English, which do not display this phenomenon.

(6)	French:	$[avek] + [vu] \rightarrow [avegvu]$	'with you'
	English:	$[blak] + [van] \rightarrow *[blagvan]$	'black van'

Fully-voiced obstruents, as we have stated, are said to contain an active element [L]. The Greek and French assimilations therefore represent the spreading to, or activation of, this element in the adjacent position(s); the lack of assimilation in English is because English lenis obstruents have no [L]. The Greek example involving the obstruent can readily be aligned with this. But it is evident that Greek nasals will also trigger this process, and the identification of [N] with [L] gives a unified account of these intuitively identical phenomena.

This, of course, leads ineluctably to the question of the contrastive representation of nasals and fully voiced obstruents language universally. One approach is outlined in Ritter (1996 and references therein). In her paper, the already extant mechanism of isomeric variation is invoked to interact with a five-item element menu (A, I, U, H and L) to express subsegmental contrasts. Broadly, headed expressions imply stricture in this system and unheaded expressions do not. So both nasals and stops are headed, but with

inverse positioning of elements, and fricatives are non-headed expressions having both place and laryngeal dependents, as illustrated in (7).

(7)	/t/	/n/	/s/
	A H	L A	(_) A H

Laryngeal contrasts within obstruents are expressed by varying the dependent element, with the neutral series lacking a dependent entirely, as in (8) (Ritter, personal correspondence).

(8)	/p ^h /	/p/	/b/	
	U H	<u>U</u>	<u>U</u> L	

In a language with fully voiced obstruents, such as modern Greek, then, we can deduce the representation of /v/ to be [U, L, (_)], and so support our account of the spread (or activation) of voicing triggered by both fricatives and nasals as identical processes.

All the discussion in this section, though drawn from separate sources, has converged upon the idea that elements, even if possessed of an identifiable singular nature as abstract cognitive objects, may adopt different shapes when linked to different structural architecture. So we may sometimes need to disentangle an element from its prosodic web to see if it really is what we think it is. The proposal that the *interaction* of prosody and melody can affect phonetic expression, just as isomeric or licensing relationships can, may prove useful in accounting for the phonological anomalies that we noticed in the analyses of lexical tone (section 3 above).

4 An experiment with tone

Returning to our discussion of lexical tone, we recall that we have so far noted the highly undesirable asymmetry of representing some west African languages as having [H] and [L] as componential in Nuclear positions and Bantu languages as having [H] only, if we

cannot discover some languages which use [L] only. One way to open enquiries into this matter would be to try to ascertain if the perceptions of native speakers of these languages accord with analyses of this type. A necessary prerequisite of such an investigation is to gain a knowledge of the perceptual routine(s) used to map from the acoustic signal to the perception of lexical tone.

Lexical tones are acoustically cued by changes in the frequency of the fundamental of the speech signal (Fx). Now the perceptions of different aspects of the speech signal are undertaken in different ways. We know that the perception of stop contrasts is quintessentially categorical, but that the strategy employed for distinguishing vowels may well be prototypical in nature (see especially Kuhl 1992). Is categorical or prototypical perception used in the perception of tone, or is such perception, as seems often to be tacitly assumed, wholly dependent upon the pitch *relationships* within an utterance?

The psychoacoustic percept of pitch is mapped from the highest common factor of the component pure tones (sine waves) of the complex wave. So this percept depends on repetition frequency, exactly as it does for a single sine wave. Just-noticeable-difference (JND) experiments in the perception of pure tone regularly show that a JND in repetition frequency of around 4Hz is perceivable at around 250 Hz, a frequency which lies squarely within the range of Fx common in human speech. Perception becomes less fine at higher frequencies, but a JND of 4Hz gives a sensible general benchmark of the physiological ability of the organism to discern pitch differences within the range normally utilised in the perception of Fx, since Hz measurements are physical, in that they are directly derivable from measurements of the linear values of air pressure in Pascals.

So we can identify a norm for the perception of pitch differentials independent of any structural context. Innately given abilities of this kind are adapted by the mind-brain when they are integrated to higher structural levels: we know of the existence of categorical perception, for instance, in animals other than *homo sapiens*. We also know that most human languages employ this ability to categorise stop consonants, but that they have no bias toward the psychoacoustic factory settings. So speakers of English and German using the lag VOT boundary (around +30ms.) share this boundary (which has high acoustic salience) with neonates (Eimas *et al.* 1971) and chinchilla (Kuhl & Miller 1975), but south of an isogloss running across northern Europe, speakers of French, Spanish, Portuguese, Greek, Dutch, Italian, and Hungarian ignore the +30ms demarcation line and instead place the boundary at the acoustically impoverished value of 0ms.

In order to discover how the perception of pitch may be lexically utilised, and how such utilisation may alter the perception of pitch, we may begin by investigating the discrimination of lexical items outside of a prosodic context by a native speaker of a tone language. Yoruba has the phonologically felicitous property of possessing many words which differ only in their tonal specification. Minimal triplets exist, such as:

(9)	bá	'meet'	kí	'greet'	rí	'see'
	ba	'hide'	ki	'thick'	ri	'weeping'
	bà	'perch'	kì	'praise'	rì	'sink/sag'

Having analysed the acoustic makeup of one of these sets, it is possible to synthesise several tokens of the words which differ only in their fundamental frequency, play them to speakers of the language, and investigate their reactions. Although productions of these words outside of a syntactic context are never going to be entirely natural, they do at least have the dual benefit of being completely linguistically plausible and contrasting solely on the target parameter, though of course more than one acoustic cue may ultimately be found to correlate with a single phonological object. Using the /ki/ set (see (9) above), I have carried out such a test, and I present the methodology and results in the Appendix.

My informant's intuitions were that it was perfectly well possible to distinguish words of Yoruba outside of any contextual frame, and the outcome of the tests, as we will see, appear to show that he is part right and part wrong.

We find that, for this particular speaker, high tone has a mean correlate in fundamental frequency of 227 Hz. That for mid tone is around 194 Hz, and for low tone 149 Hz. (These are at the *start* of the vowel: for more detail, see Appendix.) The synthetic tokens, which were replayed to the informant, a second (adult female) native speaker, and to an English-speaking control, had values ranging from 145 to 225 Hz at 5 Hz intervals. The subjects heard the tokens in pairs, each pair having a 5 to 40 Hz pitch differential. They were asked to report whether the members of the pairs were the same or different. Following this, the Yoruba speakers were asked to provide English translations for the pairs. A summary of the result for the Same/Different test is shown in (10).

(10)

Summary of average results of Same/Different responses to pairs of Yoruba words differing only in tonal specification by two Yoruba native speakers and an English control.



As is fully predictable from the JND experiments, the English control perceived a different 'word' if the signals were more than 5Hz apart. All and only the signals separated by 5Hz were perceived as the same. For the Yoruba speakers, on the other hand, anything less than 25 Hz pitch difference was not reliably perceived as different. This can be interpreted as suggesting merely that if we use pitch in a linguistically significant way, we expect a certain distance to be maintained between different phonetic categories, and so there is certainly a relativistic aspect to this perception. However, there is also an element of absolutism to be accounted for. A signal higher in pitch than 210 Hz was always, in the translation tests, perceived as an instantiation of high tone. At less than 190 Hz a mid tone was perceived. (This did not apply to low tone, as I will shortly discuss.) In between these values lay a region in which there was some confusion, as can be seen from the results (in Appendix). Also, there is evidence that a large differential will prejudice the translation: thus a value of 200 Hz is consistently translated as 'Greet' (high tone) if paired with a 160 Hz stimulus, but 'Thick' (mid tone) if paired with one of 225 Hz.

What, then, could be the strategy for the perception of lexical tone? Neither of the two routines familiar from stop and vowel perception, categorical or prototypical, seems to entirely fit the facts. There are evidently no clear cut category boundaries comparable to those developed in the perception of VOT, but there does appear to be a fuzzy acoustic no-mans-land between the pitches that are mapped to either the phonological category [H] or to the lack of *any* phonological category (\emptyset). Once either side of this area, though, there is a whole range of possible pitches that can be equally well transduced into cognition as a particular phonological object (or lack of it), so any hunt for a prototype seems doomed. Having said that, however, I wish to propose that there may be a 'post-normalised categorical' strategy employed to decode lexical tones.

I regard normalisation as an extralinguistic low level precursor to speech processing, being the selfsame process as leads us to decide in visual processing if something is 'small' or 'far away'. This is supported by the fact that normalisation is acquired extremely early in ontogeny. Visual size constancy is already present at birth (Slater 1992), and since babies recognise their mother's voice *in utero* and so can distinguish it from any other after they are born (Fifer & Moon 1989), we have some justification in proposing that we are 'normalising' sensory input before we acquire knowledge that is specific to any cognitive structure (such as a language).

Normalisation obviously cannot play a part in categorical perception, since this is by definition invariant from speaker to speaker, and equally obviously it *must* play a part in the perception of pitch in speech processing, so an unmoderated categorical routine is in any case debarred from the perception of lexical tone. But once we have 'done' our normalisation, the results of our experiment appear to show an expectancy, on the part of the native speaker of a tone language, that a particular part of the frequency range available will be used to signal a particular phonological object, even if the boundaries of that range may be swayed by the surrounding phonetic context. Consider the results of the Same/Different test for pairs of 'words' 15 and 20 Hz apart (see also Appendix):

(11)	15Hz differential	1	2	20 Hz differential	1	2
	175 185	S	S	145 165	S	S
	180 195	S	S	150 170	S	S
	210 195	D	D	210 190	D	S
	225 210	S	S	225 205	D	D
				200 220	S	D

Same (S) or Different (D) judgements by two Yoruba native speakers for pairs of stimuli having a 15 or 20 Hz differential.

At around 200 Hz of fundamental frequency there is arguably a prejudice to perceive a change in lexical category¹. No such prejudice is evident at the extremes of the ranges tested, where even a 20 Hz differential is more likely to produce a 'Same' judgement. We need to align these findings with the consistent perception of lexical change if stimuli are

¹I acknowledge that the '200-220 S' judgement does not quite fit in with this proposal: a much larger study would be needed to fully test it.

more than 20 Hz distant, and we can do this by acknowledging that tone perception comes about as the result of a 'post-normalised categorical' strategy that takes account of phonetic relativity.

No mention has yet been made in the present discussion of the most striking result of the translation tests: there is not a single instance, anywhere in the frequency range, of the perception of a low tone. Outside of a prosodic frame, the phonological object that has low tone as its corollary so far appears to be non-existent. I propose that this result be interpreted as suggesting that what we have hitherto analysed as a reflex of [L] may instead be the expression of a prosodic constituent. Naturally a prosodic constituent cannot be perceived outside of a prosodic framework. As this is a fairly radical suggestion, let us see if it has any real advantages. It is true that, on the minus side, it gives us a lot of rethinking to do with regard to the phonological analyses of the languages of western Africa. On the plus side, however, we have a reason for the lack of a single African language that can be analysed using [L] only. The story may have to be taken a whole lot further, as we now have another gap in the paradigm: we predict a language using only prosodic constituents. On the theoretical side, the proposal does tie in with the drive to reduce the combinatorial possibilities of the model (see section 3.3 above), which is mathematically desirable. We should expect to use the full set of combinations predicted in a given system, and this kind of proposition cuts down on overgeneration.

This proposal does not relieve us of the burden of explaining why [L] should be disfavoured in Nuclear positions worldwide, but if it were properly supported, would at least remove the undesirable asymmetry on which we are currently focusing².

The results of these tests, though, are far from conclusive. There are no controls for the effects of presenting the words *in pairs*, for any bias resulting from the intervention of a rise or a fall between the members of the pair, or for any change in perception that may be brought about by altering the internal shape of the pitch contour. In addition, it has been pointed out to me that the lack of any low tone perception may be the result of synthesising all the stimuli from one 'master': there could be undetected acoustic properties present in the natural low tone utterances which cue the perception of [L]³. It is also important to carry out some translation tests on single words (perhaps separated

²I will return to the 'worldwide' anomaly later.

³John Coleman pointed this out to me. I have since produced synthetic tokens of words having greater pitch variation and 'breathiness' which have yet to be tested. John Harris has proposed that it may be necessary to synthesise tokens of differing pitch from *all three* natural categories.

by music), and it could be interesting to have subjects rate the quality of various pitches as instantiations of tone(s).

5 Infant perception

The motivation for this experiment was to ascertain the perceptual routine for lexical tone so that we may see if the same routine exists in infancy, and thereby ascribe phonological significance to the perceptions of infants. We have emerged with the hypothesis that in adulthood, a native speaker of Yoruba has an active element [H] which can be linked to Nuclear positions, this having as its corollary the perception of high tone. We have proposed that we may detect the presence of this element outside any prosodic context by a fuzzy category boundary in the frequency range, the absolute value of this boundary depending on normalised perception. The boundary itself may be detected by a greater than average tendency to perceive lexical differences in its vicinity.

My intention is to import these notions into infant testing in the coming months. Using the VRISD paradigm (Eilers *et al.* 1979, Kuhl 1992 etc.), we will be able to see whether or not babies can make the phonetic discriminations which relate to these proposals. Using a sample of Yoruba acquiring children and English controls, all aged six to eight months, we can test discrimination of Yoruba words with frequency differentials of 5, 10, 15, and 20 Hz.

The hypotheses are as follows:

- (i) No child will perceive a difference for a 5 Hz frequency differential.
- (ii) English acquiring children will perceive a difference for any greater frequency differential.
- (iii) Yoruba acquiring children will not perceive differences greater than 5 Hz consistently, but there will be a tendency to perceive differences as the frequency differential grows.
- (iv) In the vicinity of the normalised category boundary between high and mid tones (around 200Hz for an adult male speaker) there will be a greater tendency on the part of the Yoruba acquiring children to perceive differences.

I hope to be able to publish the results of such tests at a later date.

6 A note on tonogenesis

The implications we have drawn so far from the testing of a Yoruba speaker have not accounted for the anomalous debarment of [L] from Nuclear positions. We might remember, though, that all the strong historical evidence for the inclusion of [L] in Nuclear positions comes from *east Asian* languages. Parallel historical developments are traceable in many languages from this part of the world. Blumstein (1991 p.109) states that the voicing contrast in Chinese and Thai obstruents which obtained at an earlier stage in the development of these languages was lost and phonologically replaced by a lexical tonal contrast. Haudricourt (1954) identifies three stages in a parallel development in the history of Vietnamese: first, an atonal language with initial voicing contrasts and final consonants; later, a three-tone system which maintains the initial contrast but has lost the final distinctive consonants. These stages can be exemplified as follows:

(12)	Stage I	pa ba	pah bah	pa? ba?
	Stage II	pa ba	paî baî	pa` ba`
	Stage III	pá bà	pấ pà	pấ pà

Of course such data are open to all kinds of *phonetic* interpretation, which can give rise to various different phonological analyses. But with that *caveat* in mind, we may be forgiven for speculating that one phonological analysis that is at least no worse than any other is that the progression between stages I and II is the transference of the [H] / Ø contrast to the Nuclear position, and that between stages II and III represents the same thing for the [L] / Ø contrast. (The orthography of these representations used here may be slightly misleading: note that in each case phonetic interpretations are substituted, rather than elided.)

So it is not surprising from the point of view of tonogenesis that there are no African tone languages utilising Nuclear [L]: a more coordinated analysis of the world's languages may in time reveal universal patterns attributable to parametrisation, and also show somewhere in the world the presence of the predicted but still unattested group of languages that use [L], and only [L], in Nuclear positions.

Appendix: methodology and results

Recording and analysis of stimuli

Recordings were made of natural utterances of three words in the Yoruba language which differ only in tonal specification, *viz*.

kí 'greet' ki 'thick' kì 'praise'

They were spoken by an adult male native speaker of the language. The stimuli were recorded in an anechoic chamber at University College London, using a Bruel and Kjaer sound level meter type 2231 with a 4165 microphone. They were transduced directly to digital audio tape using a Sony 1000ES tape recorder and then transferred to files on a Sun computer.

Ten tokens of each word were recorded and analysed using a program known as 'Esfilt'⁴ which allows audio reproduction, together with a visual display of both the amplitude signal and the spectrograph. Durations for all thirty tokens varied minimally (between about 205 and 220 ms), and there was also some variation in the amplitude displacements, particularly during release of the plosive, but such variations were only apparent across all the tokens, rather than between the sets of ten.

The pitch variation, however, clearly demarcated one word from another. The average values (in Hz) for each word of the fundamental frequencies at the start of the vowel were as follows:

kí	'greet'	227
ki	'thick'	194
kì	'praise'	149

The productions of none of the words departed more than about 3Hz from these average values. All showed a fall in fundamental frequency of 30-40Hz toward the end of the signal after maintaining a flat pitch contour for 100+ ms.

Synthesis and presentation

Of these thirty tokens, one was chosen as a model from which to copy synthesise an artificial stimulus which could subsequently be varied for a single parameter. The chosen stimulus was at the centre of the range of amplitude displacements represented in the sample and also had both a stereotypical duration (212 ms) and fundamental frequency (200 Hz at onset of voicing, 200 Hz at +160 ms from the initial burst, and 167 Hz at the last repetition of the period).

⁴Developed by Mark Huckvale at UCL.

This signal was imported into the KPE80a parallel synthesizer⁵ which allows manipulation of six formants for frequency, amplitude and bandwidth, together with frication (non-periodic excitation) and other parameters to create a copy of a natural utterance, using both spectral and spectrographic displays.

The synthesis of the artificial stimulus allowed the manipulation of the fundamental frequency independently of any other parameter to create seventeen tokens having fundamental frequencies at the onset of the vowel of between 145 and 225 Hz, each variant being 5Hz apart. The drop in fundamental towards the end of the word was retained in all cases.

These seventeen tokens were recopied to DAT tape in pairs, each pair being separated in pitch by from a minimum of 5 to a maximum of 40 Hz. There were a total of 32 pairs, each Hz differential being represented by four pairs, except the 20Hz - 5 pairs - and the 40Hz - 3 pairs. 15 of the pairs rose in pitch and 17 fell. These pairs were randomly distributed across the pitch range, and presented (on headphones) in a random order to three subjects, two Yoruba speakers (who also speak English) and one English control. They were asked to say whether the pair represented one or two words. The Yoruba speakers were then asked to provide English translations for the words on a second replay of the stimuli. The test(s) on the first Yoruba speaker were repeated one month later.

Results

a) Same/Different test

The English control recognised all and only the four pairs separated by 5Hz as 'the same'.

The Yoruba speakers recognised any pair separated by 25 + Hz as 'different'.

The Yoruba speakers recognised any pair separated by 5 or 10 Hz as 'the same' with a single exception: for speaker 2 the pair 200 190 produced a 'different' response.

The results for Yoruba speakers 1 and 2 for the 15 and 20 Hz pairs were as follows:

(11)	15Hz differential	1	2	20 Hz differential	1	2
	175 185	S	S	145 165	S	S
	180 195	S	S	150 170	S	S
	210 195	D	D	210 190	D	S
	225 210	S	S	225 205	D	D
				200 220	S	D

On repeating the test a month later with speaker 1, (only) the following differences from his original results were observed.

- i) The 225 205 pair was perceived as 'the same'.
- ii) The pair 225 200 (25 Hz differential) was perceived as 'the same'.

⁵Developed by Andrew Simpson at UCL.

	'Greet'	'Thick'	'Praise'	TOTAL
	(high tone)	(mid tone)	(low tone)	
A: 1st. test, spkr. 1	24	38	0	62
B: 2nd. test, spkr. 1	26	38	0	64
C: Speaker 2	25	39	0	64

b) Translation test: numbers of tokens of each word perceived

The discrepancy in totals between the first two tests was caused by a pair 200-190, which the subject could make no decision about the first time.

c) All tests

Every stimulus above 210 Hz was translated as 'Greet' (16 in all). Every stimulus below 190 Hz was translated as 'Thick' (24 in all).

For the other stimuli, the results were as follows:

Hz	190		195		200		205		210	
Translation	High	Mid	High	Mid	High	Mid	High	Mid	High	Mid
Test A	1	2	0	4	4	1	0	5	3	2
Test B	2	2	0	4	5	1	0	5	3	2
Test C	1	3	3	1	4	2	2	3	3	2

On tests A and B, for every pair identified as 'Same', two identical translations were provided; for every pair identified as 'Different', two different translations were provided.

On test C, there were 3 deviations in this parameter (9%).

References

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

Cabrera-Abreu, M. (1996). A phonological model for intonation without low tone. PhD dissertation. University College, University of London.

Coleman, J. & Local, J.K. (1992). The 'No Crossing Constraint' in autosegmental phonology. *Linguistics* and Philosophy 14. 295-338.

- Dell, G.S. & Reich, P.A. (1981). Stages in sentence production: an analysis of speech error data. *Journal* of verbal learning and verbal behavior 20. 611-629.
- 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.
- Eimas, P., Siqueland, E., Jusczyk, P.W. & Vigorito, J. (1971). Speech perception in infants. *Science* 171. 303-318.
- Fifer, W.P. & Moon, C. (1989). Psychobiology of newborn auditory preferences. *Seminars in Perinatology* 13. 430-433.
- Goldsmith, J.A. (1990). Autosegmental and Metrical Phonology. Oxford: Blackwell.
- Harris, J. (1994). English Sound Structure. Oxford: Blackwell.
- Harrison, P.A. (1995). Phonology and infants'perceptual abilities: asking the right questions. UCL Working Papers in Linguistics 7. 463-483.
- Haudricourt, A-G. (1954). De l'origine des tons en viêtnamien. In Journal Asiatique 242. 68-82.
- Karmiloff-Smith, A. (1995). Annotation: the extraordinary cognitive journey from fetus through infancy. *Journal of Child Psychology and Psychiatry* 36. 1293-1313.
- 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. (1992). Infants' perception and representation of speech: development of a new theory. *ICSLP Proceedings*: Banff, Canada.
- Kuhl, P.K. & Miller, J.D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190. 69-72.
- Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition* 42. 1-22.
- Local, J.K.& Lodge, K. (1996). Another travesty of representation: phonological representation and phonetic interpretation of ATR harmony in Kalenjin. *York Papers in Linguistics* 17. 77-117.
- McCarthy, J.J. & Prince, A. (1986). Prosodic Morphology. Ms, University of Massachusetts.
- Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29. 143-178.
- Nasukawa, K. (1995). Melodic structure and no-constraint-ranking in Japanese verbal inflexion. *Paper* presented at the Autumn 1995 meeting of the Linguistics Association of Great Britain, University of Essex.
- Pulleyblank, D. (1986). Tone in Lexical Phonology. Dordrecht: Reidel.
- Ritter, N. (1996). The role of asymmetrical headedness in the reduction theory of elements. *Paper presented at the GLOW colloquium, University of Athens, 1996.*
- Slater, A. (1992). The visual constancies in early infancy. The Irish Journal of Psychology 13. 411-424.
- Stemberger, J.P. (1983). Speech errors and theoretical phonology: a review. Bloomington: Indiana Linguistics Club.
- Stuart-Smith, J. (1996). The representation of a tonal contrast in British Panjabi. *Paper presented at the Fourth Phonology Meeting, University of Manchester 1996.*